DEFENCE R&D DÉFENSE

On the Theory of Particle Count Detection with an Application to the Triggering of Biological Warfare Detection Systems

Eugene Yee
Defence Research Establishment Suffield
Box 4000, Stn Main
Medicine Hat, Alberta T1A 8K6

DEFENCE RESEARCH ESTABLISHMENT SUFFIELD

Technical Memorandum
DRES TM 1999-062

March 2000

National Defence Défense nationale

DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

Canada

DEFENCE RESEARCH ESTABLISHMENT SUFFIELD
RALSTON ALBERTA

TECHNICAL MEMORANDUM 1999-062

ON THE THEORY OF PARTICLE COUNT DETECTION WITH AN APPLICATION TO THE TRIGGERING OF BIOLOGICAL WARFARE DETECTION SYSTEMS

by

Eugene Yee

PCN No. 6QD11

ACKNOWLEDGEMENTS

The motivation for this paper originated from extensive conversations with Mr Pierre Cherrier who provided valuable insights on the important issues that need to be resolved in order to improve current bio-detection algorithms. Furthermore, Mr Cherrier provided the author with his particle count data obtained by the 4WARN system at the JFT 5 field trials. These contributions are gratefully acknowledged.

ABSTRACT

A new procedure is presented for the detection of a bio-target signal in aerosol particle number count data when no prior knowledge of the existence of such a signal or of its characteristics (e.g., amplitude and shape) is assumed. Unlike previous bio-target detection algorithms, the algorithm in this paper is derived rigorously by the direct application of probability theory. To address the detection problem, probability theory is used to compare two models (or hypotheses); namely, a model (M_1) that postulates the presence of the background interference only, and an alternative model (M_2) that postulates the presence of a bio-target signal in the background interference. The posterior probability for each model is calculated based on all the available prior information, and used to determine the posterior odds ratio O_{21} in favor of model M_2 over model M_1. This ratio provides a quantitative measure of the evidence for the presence of a bio-target signal in the data. The new detection algorithm has been applied to both simulated and real particle count data and found to perform well.

EXECUTIVE SUMMARY

Title: On the Theory of Particle Count Detection With an Application to the Triggering of Biological Warfare Detection Systems

Author: Eugene Yee

Introduction: A high priority of any biological warfare (BW) agent defence system is the detection of incoming biological aerosol clouds that may contain the presence of BW agents, followed by the subsequent triggering of the system required to warn the user of the presence of the BW agent cloud and to initiate particulate air sampling for identification of the agent. In this regard, the problem of the remote detection of BW agents (e.g., bacteria, viruses) which differ in some statistical sense from their immediate background clutter is of great interest. The remote detection of BW agent contamination is desirable because it provides an early warning of a possible attack, thereby enabling defensive measures to be taken such as the donning of protective equipment (e.g., masks and clothing), the establishment of reconnaissance procedures, the avoidance of contamination, and the identification of the BW agent for initiation of medical countermeasures. Indeed, this was emphasized by Brigadier General Doesberg (Commander, Joint Program Office for Biological Detection, Alexandria, VA) who in his keynote address to a 1996 Edgewood Research, Development, and Engineering Center (ERDEC) scientific conference emphasized that "the major deficiency of our biological detection system is the lack of a reliable triggering device".

The key problem that arises in the triggering of a BW agent defence system is to determine if there is an "unusual" aerosol particle count in a particular window of particle size (e.g., respirable size range) and fluorescence intensity that deviates from the expected "normal" variability of the background particle counts in the same window. The detection of this unusual particle count in the ambient background clutter is used to trigger and warn the system of the putative presence of a possible bio-target. In almost all cases, it is necessary to detect the bio-target with no prior knowledge of its amplitude and shape and this difficulty is further accentuated because of the presence of additive background clutter which can vary in both time and space. Although the technical problem of formulating bio-target detection algorithms has been well recognized and studied over the past decade,

no good and definitive solutions have been forthcoming. One reason for this is that different investigators have appealed to their differing intuitions about how the measured aerosol particle count data from a BW detection system should be analyzed. However, intuition can give us only bits and pieces of the truth; but it almost never gives us the whole truth.

In this paper, an optimal BW detection algorithm is **derived** rigorously from connected theoretical principles, rather than formulated intuitively based on more or less *ad hoc* notions. In particular, we go back to fundamentals and consider the problem of BW detection from the start as one of probabilistic inference. Hence, in this work, we seek to understand what probability theory has to say about the problem of bio-target signal detection. The reason for this is that probability theory will give us its final verdict on the presence of a bio-target signal in the measured particle count data in the form of a single number; namely, the probability that the putative bio-target is present. In calculating it, probability theory will automatically take into account all the information in the measured particle count data that is relevant to this question, and whatever prior information is available. In so doing, the tactics that we think might succeed in the BW detection problem will turn out to be quite different from those that have been tried in the past.

Results: A new procedure is presented for the detection of a bio-target signal in aerosol particle number count data when no prior knowledge of the existence of such a signal or of its characteristics (e.g., amplitude and shape) is assumed. Unlike previous bio-target detection algorithms, the algorithm in this paper is derived rigorously by the direct application of probability theory. To address the detection problem, probability theory is used to compare two models (or hypotheses); namely, a model (M_1) that postulates the presence of the background interference only, and an alternative model (M_2) that postulates the presence of a bio-target signal in the background interference. The posterior probability for each model is calculated based on all the available prior information, and used to determine the posterior odds ratio O_{21} in favor of model M_2 over model M_1. This ratio provides a quantitative measure of the evidence for the presence of a bio-target signal in the data. The new detection algorithm has been applied to both simulated and real particle count data and found to perform well.

Significance of Results: An important advance has been made towards solving the long-standing problem concerning the detection of an unknown bio-target signal in aerosol particle count data that are contaminated by background interference events. In this problem, it is often very difficult to tell if the bio-target signal has been detected. In this paper, we present a rigorous solution to this problem that uses probability theory to derive the optimal detection procedure that needs to be applied to the particle count data. Following this process, we provide a quantitative measure based on the posterior odds ratio that answers explicitly the following question: "What is the evidence for the presence of a bio-target signal in the data?". It should be stressed that intuition alone (which in the past has been used to develop bio-detection algorithms) would never have been sufficient to suggest that the algorithm derived here from probability theory is the proper thing to do in order to optimally discriminate between background interference and a bio-target signal.

Future Work: Regarding the practical use of the proposed detection algorithm, 1200 frames of particle count data corresponding to 3600 s of sampling time currently requires about 55 s to process on a Pentium II class computer with a clock speed of 400 MHz. This is roughly a computational load of about 0.05 s per data frame, implying that the algorithm can be applied to particle count data in real time as it is acquired. Consequently, future work should focus on incorporating the new detection algorithm into a bio-detection system (e.g., 4WARN) to permit a real-time, autonomous operation of the algorithm.

TABLE OF CONTENTS

I. INTRODUCTION

A high priority of any biological warfare (BW) agent defence system is the detection of incoming biological aerosol clouds that may contain the presence of BW agents, followed by the subsequent triggering of the system required to warn the user of the presence of the BW agent cloud and to initiate particulate air sampling for identification of the agent. In this regard, the problem of the remote detection of BW agents (e.g., bacteria, viruses) which differ in some statistical sense from their immediate background clutter is of great interest. The remote detection of BW agent contamination is desirable because it provides an early warning of a possible attack, thereby enabling defensive measures to be taken such as the donning of protective equipment (e.g., masks and clothing), the establishment of reconnaissance procedures, the avoidance of contamination, and the identification of the BW agent for initiation of medical countermeasures. Indeed, this was emphasized by Brigadier General Doesberg (Commander, Joint Program Office for Biological Detection, Alexandria, VA) who in his keynote address to a 1996 Edgewood Research, Development, and Engineering Center (ERDEC) scientific conference emphasized that "the major deficiency of our biological detection system is the lack of a reliable triggering device" (as cited in [1]).

The key problem that arises in the triggering of a BW agent defence system is to determine if there is an "unusual" aerosol particle count in a particular window of particle size (e.g., respirable size range) and fluorescence intensity that deviates from the expected "normal" variability of the background particle counts in the same window. The detection of this unusual particle count in the ambient background clutter is used to trigger and warn the system of the putative presence of a possible bio-target. In almost all cases, it is necessary to detect the bio-target with no prior knowledge of its amplitude and shape and this difficulty is further accentuated because of the presence of additive background clutter which can vary in both time and space. Although the technical problem of formulating bio-target detection algorithms has been well recognized and studied over the past decade, no good and definitive solutions have been forthcoming. One reason for this is that different investigators have appealed to their differing intuitions about how the measured aerosol particle count data from a BW detection system should be analyzed (e.g, eighteen *ad hoc* BW detection algorithms have been formulated in [1] based primarily on intuitive notions). However, intuition can give us only bits and pieces of the truth; but it almost never gives

us the whole truth.

In this paper, an optimal BW detection algorithm is **derived** rigorously from connected theoretical principles, rather than formulated intuitively based on more or less *ad hoc* notions. In particular, we go back to fundamentals and consider the problem of BW detection from the start as one of probabilistic inference. Hence, in this work, we seek to understand what probability theory has to say about the problem of bio-target signal detection. The reason for this is that probability theory will give us its final verdict on the presence of a bio-target signal in the measured particle count data in the form of a single number; namely, the probability that the putative bio-target is present. In calculating it, probability theory will automatically take into account all the information in the measured particle count data that is relevant to this question, and whatever prior information is available. In so doing, the tactics that we think might succeed in the BW detection problem will turn out to be quite different from those that have been tried in the past.

In the "classical" or "orthodox" view of detection theory (or, equivalently, significance testing as it is referred to by the statisticians), the focus is on the creation of some function of the observable random variables (e.g., particle count data in our case). This function is a detection statistic, $\mathcal{L}$, that can be used to determine whether a signal is or is not present in the received data sequence. Since the detection statistic is a function of random variables, its probability distribution, assuming the truth of the hypothesis of interest (e.g., "only background interference is present in the received data" or "bio-target signal and background interference is present in received data"), can be calculated. A hypothesis is then assessed by comparing the observed value of the detection statistic with the long-run distribution of the values of the statistic in hypothetical repetitions of the measurements. Once the detection statistic has been selected and its probability distribution determined [and, frequently, even this is not possible with the result that one appeals to asymptotic (large-sample size) approximations for the determination of the statistical characteristics of $\mathcal{L}$], it must be decided how this probability distribution will be used to assess the hypothesis. A common procedure here is to use the Neyman-Pearson criterion [2], whereby the probability of false alarm is constrained to an "acceptable" value and, under this constraint, the decision rule used to determine the detection threshold is chosen to maximize the probability of detection.

This "frequentist" viewpoint of detection theory is standard and, in it, one is natu-

rally obliged to judge any detection method by its performance "in the long run", i.e., by the sampling distribution of the detection statistic when the procedure is repeated many times (and, assessed in terms of probability of detection versus probability of false-alarm). Hence, the "classical" detection theory does not refer to the performance on an *individual case* when a detection decision is made but, rather, makes statements only about long-run relative frequencies of performance and, as such, is based not only on the probability of the observed value of the detection statistic, but also on the probability of values that have not been observed as well (viz., the detection based on $\mathcal{L}$ relies on the calculation of the probability P that $\mathcal{L}$ values equal to or larger than that actually observed would be seen). Jeffreys [3] raised precisely this issue with eloquence: "*What the use of P implies, therefore, is that a hypothesis may be true may be rejected because it has not predicted observable results that have not occurred.* This seems a remarkable procedure. On the face of it, the fact that such results have not occurred might more reasonably be taken as evidence for the law, not against it." The unusual logic of P-values used in "classical" detection theory seems counter-productive since the real job before us is to make the best decision concerning the presence or absence of the bio-target signal from the information we have *in each individual case* (viz., for each received data sequence).

To see that long-run detection performance is a lesser problem, note that even if we had found a procedure whose long-run performance is proved to be as good as can be obtained for a particular ensemble (for example achieves the maximum probability of detection for a given false-alarm probability), that would not imply that this detection procedure is best—or, even tolerably good—in any particular case (i.e., for any particular received data sequence). One can trade off better detection performance for one class of samples against worse detection performance for another in a way that has no effect on long-run performance, but has a very large effect on performance in the individual case. Furthermore, for the particular problem we are interested in (i.e., detection of a bio-target signal from a dispersing BW agent cloud in received particle count data), there is not even a well-defined ensemble (e.g., the bio-target signal possesses no known form in the sense that it could vary greatly from one realization to the next, and the background interference that is determined by the environmental conditions will also necessarily vary over time). Hence, it cannot even be determined over what ensemble the long-run detection performance should be designed for, since such an ensemble is not well-defined in our problem.

It is argued here that the key activity in the bio-target detection problem is not

to follow a rule which would prove correct 90% of the time in the long run for a specific ensemble; there are an infinite number of different rules, all with this property. Rather, the key activity in this problem is draw a detection decision that is most likely to be right for the specific instance or case at hand. After all, if we have found the detection procedure which is "best" in the individual case, it is hard to see how it could fail to be "best" also in the long run. To undertake this program, we must reject all methods for detection that are predicated on the formulation of rules which give good results "in the long run" or "on the average". On an intuitive level, it is clear how to organize the reasoning in order to achieve the goal of an optimal decision for the individual case. This reasoning decomposes into the following stages: (1) try to forsee all the possibilities that might arise; (2) judge how likely each is, based on everything you see and all your past experience (prior information); (3) in light of this, judge what the probable consequences of various actions would be; and, (4) make your decision. This process of plausible reasoning preceding decisions is used in modern jurisprudence, in medical diagnosis, etc. In what follows, we will show how probability theory can be used to design a detection procedure for the bio-target detection problem that permits optimal detection decisions to be made for the individual case (rather than "for the long run"), which embodies quantitatively the stages of reasoning enunciated above.

II. THE RULES OF PROBABILITY THEORY

There are only two basic rules for manipulating probabilities; namely, the sum rule and the product rule. Because all other rules may be derived from these, the sum and product rules may be interpreted as playing the role of a "grammar" for probability theory.

In probability theory, the plausibility or probability of a hypothesis (proposition) H is assessed in light of some observed data D (more precisely, D is the proposition asserting the values of the data actually observed) and any prior information (proposition) I we may have regarding the hypothesis and the data. We write such a probability as $p(H|DI)$, i.e. the probability that "H is true given that both D and I are true". Hence, $p(H|DI)$ signifies the real number that is assigned to the probability of H given D and I. The basic rules for manipulating probabilities are the sum rule

$$p(H + \overline{H}|I) = p(H|I) + p(\overline{H}|I) = 1, \tag{1}$$

and the product rule

$$p(HD|I) = p(H|I)p(D|HI). \tag{2}$$

All the symbols appearing as arguments in Equations (1) and (2) should be understood as propositions. The symbol $\overline{H}$ denotes the negation of H (a proposition that is true if H is false). The proposition HD signifies the logical conjunction of H and D (a proposition that is true only if H and D are both true). Similarly, the proposition $H + \overline{H}$ signifies the logical disjunction of H and $\overline{H}$ (a proposition that is true if either H or $\overline{H}$ is true). The notation "$|I$)" means conditional on the truth of proposition I. Note that the plus sign appearing inside $p(H + \overline{H}|I)$ on the left-hand side of Equation (1) refers to the logical disjunction of two propositions, whereas the plus sign appearing on the right-hand side of Equation (1) refers to the addition of two numbers (i.e., the probabilities of hypotheses H and $\overline{H}$ given the prior information I). The rules in Equations (1) and (2) hold for any arbitrary propositions; and, in particular, D and I do not necessarily have to specify propositions relating to data or prior information, respectively, although this is the form we will apply these rules in this paper. Finally, we emphasize that all the probabilities in Equations (1) and (2) are *conditional*. To use the notation $p(H)$ to stand for the probability of H does not make sense until the evidence on which it is based is given (viz., knowledge is necessarily contextual). Just as the concepts of absolute space and time do not make much sense, the notion of an absolute probability $p(H)$ is meaningless because all knowledge is conditional.

The proposition "H and D" is the same as "D and H" so the truth of the propositions must be the same in the product rule. Consequently,

$$p(HD|I) = p(DH|I) = p(D|I)p(H|DI). \tag{3}$$

Equation (3) may be combined with Equation (2) to obtain Bayes rule

$$p(H|DI) = p(H|I)\frac{p(D|HI)}{p(D|I)}. \tag{4}$$

Bayes rule describes a type of learning: how the probability of a hypothesis should be modified on obtaining new information D in the form of observed data. The probability $p(H|I)$ for the hypothesis H in the absence of D is called the prior probability, and the probability $p(H|DI)$ including the information D is called the posterior probability. The quantity $p(D|HI)$ is called the sampling probability for D (when H is fixed), or the likelihood for

H (when D is fixed). The quantity $p(D|I)$ is called the prior predictive probability for D, and essentially plays the role of a normalization constant in Equation (4). Bayes rule is the starting point for all inference problems using probability theory. As it currently stands, Bayes rule is not sufficient to conduct inference because, ultimately, the numerical values of the probabilities must be known to begin a calculation. How this task needs to be accomplished for the bio-detection problem will be described in the next two sections.

III. CHARACTERIZATION OF BACKGROUND PARTICLE COUNTS

The detection of the bio-target signal must be undertaken in the presence of a background particle count interference associated with sulfates, nitrates, pollen, fungi, spores, and other naturally occurring organisms and organic matter that is present in the atmosphere. This "noise" can also include measurement noise in the detection system such as background optical noise arising from the effects of stray laser light reaching the fluorescence photomultiplier tube ("dark current"). Whatever the source for the background interference within a prescribed window of particle size and fluorescence intensity, the detection algorithm must be able to estimate the total background noise in the data and make proper allowances for it. Consequently, we shall assume that the level of the background particle counts and its variability in a prescribed window of particle size and fluorescence intensity are not known in advance, so the detection algorithm must estimate this information from the data and use it to determine the reliability or accuracy of the subsequent bio-target detection. This is important because the presence of the intrinsic background particle counts will degrade the sensitivity of the detection because it is impossible to unambiguously separate events originating from the bio-target signal process from the expected background events.

To this end, we assume that we have available an independent sample of the background interference (i.e., sample that has been independently measured). The background interference data are the counts of aerosol particles in a prescribed window of particle size and fluorescence intensity when it is known *a priori* that there is no bio-target signal present. Suppose that this background interference measurement yielded n_b counts in an interval of duration T. Since we are dealing with measurements that are obtained by counting particles, the Poisson distribution is the appropriate sampling distribution (or, equivalently, likelihood function) in this case [4]. Hence, the background interference can be characterized

by an unknown expected mean b (background rate); viz., we make the hypothesis H that the background interference can be characterized by a Poisson process with background rate b. We will assume that the background rate b does not vary in the sampling interval T. However, this does not preclude the case of a slowly time-varying background rate over a time scale $T^* \gg T$ [viz., each background reception of sampling interval T used as the reference for the next (or next series) of detection decisions can have a different b, whose value will be estimated anew for each reception]. Hence, b can vary with time, but on a time scale T^* that is long compared to the sampling time for the background reception.

The value of b for each background reception can be estimated from the background interference data by simply taking the prior information I_b as specifying the connection between b, n_b, and T. Consequently, I_b will identify the Poisson distribution as the likelihood function; viz., the probability of seeing n_b background events in an interval T (which we identify with the data hypothesis D, so $D \equiv n_b$) assuming a background rate b is

$$p(n_b|bI_b) = \frac{(bT)^{n_b}\exp(-bT)}{n_b!}. \tag{5}$$

It will be assumed that we know nothing about the value of b before the receipt of any of the background samples. The least informative prior for the rate of a Poisson distribution can be derived from a simple group invariance argument, noting that b^{-1} plays the role of a scale for the measurement of time [5]. This argument results in the assignment of the Jeffreys prior [3] as the prior distribution for b (more, precisely, for the hypothesis H that the background events are characterized by the rate b):

$$p(b|I_b) = \frac{1}{b}. \tag{6}$$

In essence, this corresponds to a prior that is uniform in $\log(b)$, and expresses complete ignorance regarding the scale of the background rate (viz., we do not know anything about the background rate before the measurements of the background). Finally, the prior predictive probability $p(n_b|I_b)$ can be obtained from Equation (5) by integrating over all possible values of b, so

$$\begin{aligned} p(n_b|I_b) &= \int_0^\infty p(n_b|bI_b)\,db \\ &= \int_0^\infty \frac{(bT)^{n_b}\exp(-bT)}{n_b!}\,db \\ &= \frac{1}{n_b}. \end{aligned} \tag{7}$$

Given the probability distributions in Equations (5), (6), and (7), Bayes rule of Equation (4) can be applied to calculate the posterior distribution for the background rate b (which encapsulates our state of knowledge of b after the background measurement):

$$\begin{aligned} p(b|n_b I_b) &= p(b|I_b)\frac{p(n_b|bI_b)}{p(n_b|I_b)} \\ &= \frac{1}{b}\cdot\frac{(bT)^{n_b}\exp(-bT)}{n_b!}\cdot n_b \\ &= \frac{T(bT)^{n_b-1}\exp(-bT)}{(n_b-1)!}. \end{aligned} \tag{8}$$

Note that the posterior distribution for the (unknown) background rate is a gamma distribution with mean value $\langle b\rangle = n_b/T$ and standard deviation $\sigma_b = n_b^{1/2}/T$. The latter is the usual "root N" result expected from a Poisson signal. The posterior distribution for the background rate b in Equation (8) encodes all the information we have about the background interference process in our assumed current state of knowledge (viz., in the state of knowledge where we have a sample of the background interference consisting of n_b counts obtained in a time T). This distribution is the full Bayesian solution to the problem of estimating b. This full result will be used in the bio-detection problem considered in the next section.

For some applications, it may be useful to simply summarize the information embodied in Equation (8) in the form of a Bayesian interval $[b_1, b_2]$ corresponding to a confidence level α which can be constructed as follows:

$$\int_{b_1}^{b_2} p(b|n_b I_b)\, db = \alpha. \tag{9}$$

These intervals are probably more properly described as "credible intervals"; viz., an allowed range for b (background rate) with probability content α. Note that there is freedom in the choice of b_1, depending on whether one desires an upper credible, lower credible, or central credible interval. An alternative summary of the information encoded in Equation (8) about b may be to simply report the posterior mean, $\langle b\rangle$, and the "one sigma" credible region of $\langle b\rangle \pm \sigma_b$ (viz., $(b)_{\text{est}} = \langle b\rangle \pm \sigma_b$ which is simply a "best-fit" value and "error bars"). However, the definition of a credible region for b (either an α confidence level or a "one sigma" level) should be viewed as indices that describe only specific characteristic points about the background interference. All the information in the measured background sample is contained in the posterior distribution for b exhibited in Equation (8).

IV. DERIVATION OF DETECTION ALGORITHM

Now, let us suppose that we are given a new measurement. In particular, we are told n particle counts in a sampling time Δt have been observed during an interval when there is a suspected bio-target signal. Given the information on the background interference (cf. Section III), we need to decide between two possibilities (hypotheses); namely, either that the new measurement is consistent with the background interference, or that there is a bio-target signal present in addition to the background interference. In many cases, it is often very difficult to tell if the bio-target signal has been detected. The rules of probability theory outlined in Section II can be used for this purpose, but to do so one must state exactly what one means by "signal" and by "no signal". In this problem, "no signal" will be represented by model M_1

$$r = b, \tag{10}$$

and the bio-target "signal" will be represented by model M_2

$$r = s + b, \qquad s > 0, \tag{11}$$

where r is the (unknown) event rate corresponding to the new measurement, b is the (unknown) background event rate, and s is the (unknown) bio-target signal event rate. In Equation (10), the exact value of the background event rate b is unknown, although the assumed availability of a sample of the counts from the background imposes constraints on its probable values as derived in Section III. Furthermore, when the bio-target signal is present as in Equation (11), the exact value of the signal event rate in the given sampling interval is assumed not to be known *a priori*. However, given that the bio-target was present when the new measurement was made, this measurement must embody information about both b and s.

Given the two models, Equations (10) and (11), the problem of bio-target detection reduces to the following question: "What is the evidence in favor of model 2 over model 1?". To address this problem, we need to calculate the posterior probability of each model (i.e., either M_1 or M_2), conditioned on the new data D and any prior information I. In this case, we take as our prior information the proposition that one of the two given models is true, i.e., $I = M_1 + M_2$ where the plus sign here stands for logical disjunction. Bayes rule implies that the posterior probability of model M_i $(i = 1, 2)$ is given by

$$p(M_i|DI) = p(M_i|I)\frac{p(D|M_iI)}{p(D|I)}, \qquad i = 1, 2. \tag{12}$$

Once these two posterior probabilities have been evaluated, it is useful to consider their ratio. The ratio, $O_{21} \equiv p(M_2|DI)/p(M_1|DI)$, which is the posterior odds ratio of model M_2 over model M_1, is given explicitly by [cf. Equation (12)]

$$O_{21} \equiv \frac{p(M_2|DI)}{p(M_1|DI)} = \frac{p(M_2|I)}{p(M_1|I)} \frac{p(D|M_2I)}{p(D|M_1I)} = \frac{p(M_2|I)}{p(M_1|I)} \frac{p(D|M_2)}{p(D|M_1)}, \tag{13}$$

where in the second line of Equation (13) we used the fact that the compound proposition M_iI is true if and only if model M_i is true; i.e., it is equivalent to the proposition M_i itself, so $p(D|M_iI) = p(D|M_i)$.

To calculate the posterior odds ratio of Equation (13), we need to compute two terms: the first, $p(M_i|I)$ $(i = 1, 2)$ is the probability of model M_i given only the prior information I. This term represents one's state of knowledge about each of the two possible models before the new data D is obtained. The second term, $p(D|M_i)$ $(i = 1, 2)$, is the global likelihood of the data D given model M_i. It represents how well the model fits the data. Note that the first factor $[p(M_2|I)/p(M_1|I)]$ in the posterior odds ratio of Equation (13) can be interpreted as the prior odds ratio, and the second factor $[p(D|M_2)/p(D|M_1)]$ can be interpreted as a Bayes factor consisting of the ratio of the global likelihoods of the two models. It is important to emphasize that the model comparison here relies on the ratio of global likelihoods, and not maximum likelihoods. An important consequence of this fact is that the marginalization used to calculate the global likelihoods (see below) implies that the Bayes factor automatically favors the simpler model (i.e., the model for background interference only) unless the data justify the complexity of the more complex alternative (i.e., the model for bio-target signal and background interference).

To proceed further, we need to assign the two terms, $p(M_i|I)$ and $p(D|M_i)$ $(i = 1, 2)$. In our current problem, it is assumed that little prior information about the presence or absence of a bio-target signal within a given time window is available *a priori*. In consequence, we will consider the hypothesis of the presence and absence of a bio-target signal to be equally probable *a priori* and assign equal probabilities of $1/2$ to each of the two models. Thus, $p(M_i|I) = 1/2$ $(i = 1, 2)$ and the prior $p(M_i|I)$ will cancel in Equation (13), leaving only the global likelihood of the data in the determination of the posterior odds ratio. However, in the case where military intelligence suggests that a BW attack is likely, then this state of knowledge would be consistent with the assignment $p(M_2|I) \equiv \mathcal{P} > p(M_1|I) \equiv$

$(1 - \mathcal{P})$. The value assigned to $\mathcal{P}$ will depend on how likely military intelligence deems the BW attack to be. The numerical values assigned to the prior probabilities $P(M_i|I)$ will depend on the scenario, but to provide a specific example in the following discussion, we will assume that we do not have any specific prior information concerning whether a BW attack is or is not imminent, so we simply assign $p(M_1|I) = p(M_2|I) = 1/2$ (Laplace's principle of indifference) [6].

The global likelihood of the data, $p(D|M_i)$, for model M_i is obtained from the joint probability of the data and the parameters of the model. For model M_1 [cf. Equation (10)], the only parameter is the background rate b, so for this model the relevant joint probability of the data and the model parameter can be expressed as $p(D, b|M_1)$. However, the exact value of b is unknown, so we need to remove this parameter from the problem. To accomplish this, we simply apply the sum rule to integrate out the nuisance parameter b:

$$\begin{aligned} p(D|M_1) &= \int_0^\infty db\; p(D, b|M_1) \\ &= \int_0^\infty db\; p(b|M_1) p(n|bM_1), \end{aligned} \tag{14}$$

where the integral is over all possible values of the parameter b (which in our case can assume a continuum of positive values). In the second line of Equation (14), the product rule [Equation (2)] was used to express the joint probability $p(D, b|M_1) \equiv p(n, b|M_1)$ in terms of the prior probability $p(n|M_1)$ and the likelihood function $p(n|bM_1)$ (recognizing that the new data D here consists of observing n counts in the sampling time Δt). The prior probability $p(b|M_1)$ for this problem is informative, since we assumed in Section III that we have some background interference data available before the measurement of D. In particular, the prior probability $p(b|M_1)$ in this problem is just the posterior probability for b from the background estimation problem in Section III, viz., $p(b|M_1) = p(b|n_b I_b)$ where $p(b|n_b I_b)$ has been determined in Equation (8). Furthermore, the likelihood $p(n|bM_1)$ here is simply the Poisson distribution for obtaining n counts in the sampling time Δt assuming a given (constant) background rate b, so

$$p(n|bM_1) = \frac{(b\Delta t)^n \exp(-b\Delta t)}{n!}. \tag{15}$$

Now, inserting Equations (8) [for $p(b|M_1)$] and (15) [for $p(n|bM_1)$] into Equation (14) and evaluating the resulting integral, the global likelihood of the data for model M_1 takes the following form:

$$p(D|M_1) = \int_0^\infty db\; \frac{T(bT)^{n_b - 1} \exp(-bT)}{(n_b - 1)!} \cdot \frac{(b\Delta t)^n \exp(-b\Delta t)}{n!}$$

$$= \frac{T^{n_b}(\Delta t)^n}{(T+\Delta t)^{n+n_b}} \frac{\Gamma(n+n_b)}{(n_b-1)!n!}, \tag{16}$$

where $\Gamma(x)$ denotes the gamma function [7].

Next, we need to evaluate the global likelihood of the data for model M_2. Model M_2 has two parameters; namely, s and b [cf. Equation (11)], both of which are unknown. It is a simple consequence of the sum and product rules that,

$$\begin{aligned} p(D|M_2) &= \int_0^\infty ds \int_0^\infty db\; p(D,s,b|M_2) \\ &= \int_0^\infty ds \int_0^\infty db\; p(s,b|M_2)p(D|sbM_2) \\ &= \int_0^\infty ds \int_0^\infty db\; p(b|M_2)p(s|bM_2)p(n|sbM_2). \end{aligned} \tag{17}$$

In the second and third lines of Equation (17), the product rule was applied to factor the joint probability densities $p(D,s,b|M_2)$ and $p(s,b|M_2)$. The statistical characteristics of the background interference measured in Section III are assumed to be identical to that contaminating the bio-target signal in model M_2. In consequence, $p(b|M_2) = p(b|n_b I_b)$, so the prior distribution for b in this evaluation is identical to the posterior distribution for b obtained from the background interference estimation problem in Section III. For the prior distribution for s given the background rate b, we assign the least informative prior (i.e., Jeffreys prior [3]) for a Poisson rate $(s+b)$, with the value of b given, viz.

$$p(s|bM_2) = \frac{1}{s+b}. \tag{18}$$

The likelihood $p(D|sbM_2) \equiv p(n|sbM_2)$ in Equation (17) is identified as the probability for obtaining n counts in a sampling time of Δt for a Poisson rate of $(s+b)$ [viz., in model M_2, the putative measurement counting rate is $(s+b)$ since the observation is assumed to contain both the bio-target signal and the background events]. Hence,

$$p(n|sbM_2) = \frac{\left((s+b)\Delta t\right)^n \exp(-(s+b)\Delta t)}{n!}. \tag{19}$$

Substituting Equations (8) [for $p(b|M_2)$], (18) [for $p(s|bM_2)$], and (19) [for $p(n|sbM_2)$] into Equation (17), we get

$$\begin{aligned} p(D|M_2) = \int_0^\infty ds \int_0^\infty db\; & \frac{T(bT)^{n_b-1}\exp(-bT)}{(n_b-1)!} \cdot \frac{1}{(s+b)} \\ & \times \frac{\left((s+b)\Delta t\right)^n \exp(-(s+b)\Delta t)}{n!}. \end{aligned} \tag{20}$$

All that remains to formally complete the problem is to evaluate the indicated integral in Equation (20). Surprisingly, this integral can be evaluated in closed form, so a numerical method for integration is not required. There are a number of different ways that the integral in Equation (20) may be done, but perhaps the easiest is to rewrite the integral in the following form:

$$p(D|M_2) = \frac{T^{n_b}(\Delta t)^n}{(n_b-1)!n!}\int_0^\infty db\; b^{n_b-1}\exp(-b(T+\Delta t))\int_0^\infty ds\;(s+b)^{n-1}\exp(-s\Delta t). \quad (21)$$

Next, we focus on evaluation of the inner integral with respect to s in Equation (21). To accomplish this, make the substitution $\kappa = s\Delta t$ so $d\kappa = \Delta t ds$, and rewrite the inner integral of Equation (21) as

$$\begin{aligned}\int_0^\infty ds(s+b)^{n-1}\exp(-s\Delta t) &= \int_0^\infty \frac{d\kappa}{\Delta t}\left(\frac{\kappa}{\Delta t}+b\right)^{n-1}\exp(-\kappa)\\ &= \frac{\exp(b\Delta t)}{(\Delta t)^n}\int_{b\Delta t}^\infty dv v^{n-1}\exp(-v)\\ &= \frac{\exp(b\Delta t)}{(\Delta t)^n}\,\Gamma(n;b\Delta t), \end{aligned} \quad (22)$$

where $\Gamma(\nu;x)$ denotes the complementary incomplete gamma function [7]. In the second line of Equation (22), we made the substitution $v = (\kappa/\Delta t + b)\Delta t$; and, in the third line of Equation (22) we used the fact that the integral in the second line is simply the definition for the complementary incomplete gamma function [7]. Now, if we insert Equation (22) into Equation (21), we get

$$p(D|M_2) = \frac{T^{n_b}}{(n_b-1)!n!}\int_0^\infty db\; b^{n_b-1}\exp(-bT)\Gamma(n;b\Delta t). \quad (23)$$

The integral in Equation (23) can be evaluated by using the fact that ([8], pg. 309)

$$\int_0^\infty x^{\mu-1}\exp(-\beta x)\Gamma(\nu;\alpha x)\,dx = \frac{\alpha^\nu\Gamma(\mu+\nu)}{\mu(\alpha+\beta)^{\mu+\nu}}\;{}_2F_1\big(1,\mu+\nu;\mu+1;\beta/(\alpha+\beta)\big), \quad (24)$$

where ${}_2F_1(a,b;c;x)$ is the hypergeometric function with numeratorial parameters a and b, denominatorial parameter c, and argument x. This finally gives

$$p(D|M_2) = \frac{T^{n_b}(\Delta t)^n}{n_b!n!}\cdot\frac{\Gamma(n+n_b)}{(T+\Delta t)^{n+n_b}}\;{}_2F_1\big(1,n+n_b;1+n_b;T/(T+\Delta t)\big). \quad (25)$$

The posterior odds ratio O_{21} for model M_2 over model M_1 can now be explicitly exhibited by substituting Equations (16) and (25) into Equation (13) and noting that for our example we have assumed that $p(M_1|I) = p(M_2|I) = 1/2$:

$$O_{21} = \frac{1}{n_b}\;{}_2F_1\big(1,n+n_b;1+n_b;T/(T+\Delta t)\big). \quad (26)$$

This is simply the odds in favor of model M_2 (viz., the model that proposes that a bio-target signal is present in the data D). The odds ratio of Equation (26) expresses a bet. If the odds ratio is greater than 1, it is bet in favor of the model postulating the bio-target signal. If the odds are less than 1, it is a bet in favor of the model proposing that only background interference is present—viz., there is no bio-target signal of interest. If the odds are exactly 1, neither model is to be preferred.

Equation (26) provides the evidence for the possible presence of a bio-target signal in the received data sequence in the form of an odds ratio. This ratio provides us with the final state of knowledge with all the available prior information (background samples) and data (new received data sequence) taken into account. All the activity described above and culminating in Equation (26) concerns inference. An essential thing which is still missing is the rule which converts the odds ratio (evidence) contained in Equation (26) into a definite course of action (decision theory). Hence, we need to decide how large the posterior odds ratio O_{21} needs to be in order to declare the presence of a bio-target signal in the data. With regard to this problem, which belongs to "decision theory" proper, there is nothing in probability theory *per se* which could tell us where to put the critical levels at which the decision for the detection of a bio-target is made: viz., whether the putative target is present or absent in the received data sequence. The location of these critical levels obviously depends in some way on value judgements as well as on probabilities; what are the consequences of making wrong decisions, and what are the costs of making further measurements before the decision is made?

To this end, we need some specific criteria of what we want our BW detection system to accomplish. The criteria will vary with the application and necessarily depend on certain value judgements, and obviously no single decision rule can be best for all purposes. A naïve choice may be to simply declare the bio-target signal to be present whenever O_{21} is greater than 1 (viz., a bio-target signal is declared whenever M_2 is preferred over M_1). A more general approach will require the assignment of a loss function $L(P, M)$ which represents our judgement of how serious it is to make decision P when model M is in fact true.

In our problem, there are only two possible models; M_1 (only background interference is present) and M_2 (bio-target signal is present in the background interference) which must be discriminated against from the given particle count data D. In consequence, there are two possible decisions P_1 and P_2 corresponding, respectively, to the absence and presence

of a bio-target signal in D. Hence, there are two types of errors; the false alarm $A = (P_2, M_1)$ (a bio-target is detected when none actually existed), and the false rest $R = (P_1, M_2)$ (a bio-target is not detected when it was actually present). In certain BW applications, one type of error might be more serious than the other. As an example, suppose that a false rest is considered ten times as serious as a false alarm, while a correct decision of either type represents no "loss". We could then take $L(P_1, M_1) = L(P_2, M_2) = 0$, $L(P_2, M_1) \equiv L_a = 1$ and $L(P_1, M_2) \equiv L_r = 10$, where L_a and L_r are the losses incurred by a false alarm and a false rest, respectively. In this scenario, a reasonable way to proceed may be to make the decision which minimizes the expected loss. This simply leads to the following decision rule:

$$\text{Choose } P_2 \text{ if } \quad O_{21} > \frac{L_a}{L_r}. \tag{27}$$

Note that the naïve choice mentioned above simply corresponds to the case where the losses incurred from a false alarm and a false rest are judged to be equally serious. It should be stressed that the assignment of a loss for a false alarm or false rest and, hence, determination of the critical levels is, of course, a question for operational analysts and military commanders to decide for a particular operational scenario and one for which physical science has no bearing.

V. EXAMPLES OF APPLICATION OF DETECTION ALGORITHM

We have shown that if our task is to make the best possible decision as to whether a bio-target signal is present, the logical thing we must do is to calculate the probability that the signal is present, conditional on all the evidence at hand. If there are only two possibilities, M_1 and M_2 (as in our problem), to be taken into account, then after we have seen the new data D, the posterior odds on M_2 is given by Equation (26), which is equal to the prior odds multiplied by a dimensionless factor which we called the global likelihood ratio (Bayes factor). In our application, it is convenient to take the logarithm of the posterior odds ratio because of the large dynamic range exhibited by this quantity. To this end, we define a new function which we will call the *evidence* for M_2 given data D and prior information I:

$$K \equiv 10 \log_{10} O_{21}. \tag{28}$$

This is still a monotonic function of probability. By using the base 10 and the factor 10 in front, we are measuring evidence in units of decibels (dB).

We now have three different scales on which we can measure degrees of plausibility for M_2 (i.e., presence of a bio-target signal) given the data D and prior information I; namely, evidence, odds, or probability. These measures are all monotonic functions of each other. Zero dB of evidence corresponds to odds of 1 to 1 or a probability of 1/2, that is, neither model (M_1 or M_2) is to be preferred. If the evidence is 20 dB, then the odds are 100 to 1 or the probability is 100/101 in favor of model M_2. Similarly, if the evidence is -20 dB, then the odds are 100 to 1 or the probability is 100/101 in favor of model M_1 (i.e., no bio-target signal is present in the data).

By experimenting with some examples, a better understanding of the detection process can be obtained. To begin, we illustrate the use of our detection method by applying it to simulated data. We perform detection calculations both for count data from a finite-duration step bio-target signal with a constant expected (mean) signal rate and count data from a transient bio-target signal whose expected signal count varies with time.

To begin, we simulated a Poisson background event count with an expected constant background rate $\langle b \rangle = 15\ \mathrm{s}^{-1}$ over an observation interval of 3600 s. These counts were further segmented into 3 s sampling intervals to simulate a BW detection system that provides a time sequence of particle number counts within a prescribed window of particle size and fluorescence intensity. Henceforth, the particle number counts in a prescribed window of particle size and fluorescence intensity for each 3 s interval will be referred to as a frame of data. Consequently, the simulated background event counts for an observing interval of 3600 s corresponds to 1200 frames of data.

For our first example, we simulated a finite duration bio-target signal with a constant expected signal rate $\langle s \rangle = 7.5\ \mathrm{s}^{-1}$ and a duration of 300 s (100 frames). This bio-target signal count was added to the background event count beginning at frame 501. This gives a signal-to-noise ratio (SNR) of 0.5 where SNR $\equiv \langle s \rangle / \langle b \rangle$. Figure 1(a) shows the resulting particle count data. Figure 1(b) is a plot of the evidence K for the data shown in Figure 1(a). The background data used for this calculation was obtained from the first 50 frames ($T =$ 150 s) of the data in Figure 1(a). The evidence in this example (and all the examples to follow) was calculated by applying the detection algorithm [Equation (26)] sequentially to the data in Figure 1(a), with the evidence at a particular frame of data obtained from the particle counts measured in a window of 7 frames centered on the particular frame. From Figure 1(b), one concludes that the data shown in Figure 1(a) contains positive evidence in

favor of a bio-target signal in frames 501 to 600 [see Figure 2(b) where the bio-target signal was declared present whenever K exceeds 0 dB]. In particular, within the time interval from frames 501 to 600, the evidence exceeds 160 dB which implies a bet of better than 10^{16} to 1 in favor of the presence of a bio-target signal.

Figure 3(a) shows simulated particle count data corresponding to the case with an expected constant signal rate $\langle s \rangle = 5.0\ \mathrm{s}^{-1}$, resulting in a signal-to-noise ratio of 0.333. The evidence in favor of Model M_2 for the data in Figure 3(a) is plotted in Figure 3(b). For these data, at the maximum, there is a positive evidence in excess of 80 dB for the presence of a bio-target signal in the time interval between frames 501 and 600. This corresponds to a bet of better than 10^8 to 1 in favor of the model containing the bio-target signal. Figure 4(b) shows the detection of the bio-target signal using a threshold of 0 dB in the evidence.

Figure 5(a) displays simulated particle count corresponding to a signal-to-noise ratio of 0.25, and Figure 5(b) shows the associated evidence for the presence of a bio-target signal. The detection results for this example are exhibited in Figure 6(b) using a threshold of 0 dB for the evidence. In spite of the low signal-to-noise ratio, the evidence in favor of the presence of the bio-target signal reached a maximum of 50 dB in the interval between frames 501 and 600, implying that the odds ratio in favor of the bio-target signal would be approximately 10^5 to 1.

The final example of a step bio-target signal is shown in Figure 7(a) for the case where the signal-to-noise ratio was 0.2. Figure 7(b) shows the corresponding evidence K, whereas the detection results are displayed in Figure 8(b) for a threshold of 0 dB in the evidence. With this threshold, the signal was detected in 20 frames in the interval between frames 501 and 600, where at the maximum, it is a bet of better than about 300 to 1 (about 25 dB in evidence) in favor of the presence of a bio-target signal.

The next simulated example considers the case of a transient bio-target signal with a time-varying expected signal rate $\langle s(t) \rangle$ given as follows [cf. Figure 9]:

$$\langle s(t) \rangle = \frac{A t^{\nu/2-1} \exp(-t/2)}{2^{\nu/2}\Gamma(\nu/2)}, \tag{30}$$

where $\nu = 10$ and the amplitude A was adjusted to give a maximum expected signal rate of about $5.0\ \mathrm{s}^{-1}$. When this signal is added to the background event counts generated

previously, this will give data with a maximum SNR of about 0.333. A realization of a transient bio-target signal count with the time-varying expected signal rate shown in Figure 9 is displayed in Figure 10(a). When this signal is added to the background interference events between frames 401 and 600, it gives the particle number counts shown in Figure 11(a). The variation of the SNR over the time interval spanning these 200 frames is exhibited in Figure 10(b). The evidence K for the presence of a bio-target signal in the data shown in Figure 11(a) is exhibited in Figure 11(b). Figure 12(b) shows the results of signal detection using a threshold of 0 dB in the evidence (implying that the losses incurred by a false alarm and a false rest are assigned equal weights). At the peak of the bio-target signal strength, probability theory finds an evidence for the presence of a signal of about 50 dB (viz., a bet of about 10^5 to 1 in favor of the presence of a bio-target signal).

Next, we present some results of the application of our detection algorithm to data obtained from the Combined Joint Field Trials For Biological Detection (JFT 5) conducted at a site Tower Grid of the West Desert Test Center Dugway Proving Ground (40°06' N, 112°59' W) in northern Utah. The site is 2 km west of Camel Back Ridge on the edge of the Great Salt Lake Desert, and is flat and relatively smooth and homogeneous, with an unobstructed fetch of more than 5 km in a wide sector. The data used for testing our detection algorithm was measured by the 4WARN real-time integrated biological and chemical agent detection system developed by Computing Devices Canada Ltd. The main component of 4WARN for BW agent detection is the Fluorescence Aerodynamic Particle Sizer (FLAPS) which can output total particle number counts in a prescribed window of aerodynamic particle size and fluorescence intensity as a function of time. The design and construction of FLAPS is described in Hairston *et al.* [9]. For the measurements in JFT 5, FLAPS was operated in 4WARN to provide a frame of particle count data every 3 s.

Figure 13(a) shows 800 frames of particle count data obtained by 4WARN for JFT 5 Trial 19. Because no independent background count data was provided for this trial, we used the particle count data measured between frames 151 and 200 inclusive [i.e., 50 frames of background data corresponding to a sampling time $T = 150$ s] as a "representative" sample of the background interference for this trial. The detection algorithm consists of computing the odds in favor of the presence of a bio-target signal in background interference model (M_2) over the background interference only model (M_1). Figure 13(b) is a plot of the evidence K (logarithm of the odds ratio) as a function of time (frame number). Between frames 500 and 600, the largest value in the evidence exceeds 300 dB, implying the odds in

favor of a bio-target signal in the data of better than 10^{30} to 1 within this interval of time. The resulting odds indicate near certainty of the presence of a bio-target signal. Figure 14(b) shows the indicator function corresponding to detection of the bio-target signal. For this example, detection of the signal was accomplished when the odds ratio exceeded 5.0 (i.e., the evidence exceeded 7 dB). In selecting this threshold for detection, it was implicitly assumed that the loss incurred by the system arising from a false alarm was 5 times as serious as the loss arising from a false rest. In Figure 14(b), the detection indicator function can be compared to the reference data obtained independently on a nearby referee tower using all-glass impingers (AGI). The reference data are provided in aerosol cloud particles per liter of air (ACPLA).

In this example, the reference data obtained from the AGI appeared to correlate well with the evidence K [cf. Figures 14(a) and (b)]. In addition to the detection of the main "core" of the cloud, the algorithm was also able to detect the presence of a long and persistent cloud concentration "tail" which persisted after the passage of the main cloud from about frames 600 to 750. The evidence for presence of a bio-target signal in this time interval varied from about 30 to 50 dB, giving odds of between 1000 and 100000 to 1. The presence of a persistent concentration tail in the cloud is confirmed in the reference data, which shows low concentrations of between 1 and 3 ACPLA over this time interval.

VI. CONCLUSIONS

In this paper, we have shown how to derive an optimal detection algorithm for particle count data, when we have no *a priori* knowledge about the nature of the bio-target signal (i.e., its amplitude, shape, and arrival time). This is accomplished by using probability theory to calculate the probability for two mutually exclusive but exhaustive hypotheses; namely, the hypotheses that "only background interference is present" and that "a bio-target signal is present in the background interference". The result is a bio-target detection algorithm that is derived from connected theoretical principles, rather than obtained from some intuitive and *ad hoc* procedure that has characterized the development of these algorithms to date.

A quantitative measure (evidence or, equivalently, the logarithm of the posterior odds ratio) has been provided which provides the explicit answer to the following question: "What is the evidence in favor of the presence of a bio-target signal in the measured

data, given all the prior information at hand (e.g., a limited sample of the background interference)?". Probability theory has provided this evidence in terms of an odds ratio (cf. Equation (26)) which represents the final state of knowledge with all the available prior information and data taken into account. This corresponds to the problem of inference. The final part of the problem requires the determination of critical levels that could be used to convert the evidence in the form of an odds ratio into a definite course of action (i.e., decide whether a bio-target is present or absent). This part of the problem corresponds to decision theory, and as such, must necessarily depend in some way on value judgements as well as probabilities. The setting of critical levels for detection will depend on the particular scenario as well as on military doctrine. In this way, probability theory and decision theory have realized quantitatively the "reasoning preceding decisions" paridigm alluded to in the introduction: namely, (1) try to forsee all the possibilities that might arise (e.g., enumeration of the relevant hypotheses such as "background interference present only" and "background interference and bio-target signal present"); (2) judge how likely each is, based on everything you can see and all your past experience (e.g., evidence is quantitatively encapsulated in odds ratio of Equation (26) which embodies prior information on the background samples and new data in the form of the received data sequence); (3) in the light of this, judge what the probable consequences of the various actions would be (e.g., formulated in terms of loss functions which encapsulate the losses that would be incurred from a "wrong" decision); (4) now make your decision (e.g., "bio-target present" or "bio-target absent").

It should be emphasized that the proposed approach for the design of bio-detection algorithms differs radically from the frequentist viewpoint of detection theory whereby the detection algorithm is judged by its performance "in the long run" (i.e., by the sampling distribution of the detection statistic when the procedure is repeated many times and assessed in terms of probability of detection versus probablity of false-alarm). In this traditional method, one considers the probabilities of hypothetical data, and assumes the truth of the best-fit parameter values. In stark contrast, the proposed method in this paper uses only the probability of the actually observed data, and additionally takes into account all the possible parameter values (e.g., s and b, which are the unknown signal and background rates) through marginalization. In our proposed approach, we are not interested in long-run detection performance, but rather with the specification of a procedure that provides the "best" performance in each individual case, where we know from the start that this case can never be repeated (either in the signal received or the background present or both).

In this regard, we note that the sampling distribution used in the frequentist viewpoint of detection theory (e.g., Neyman-Pearson theory which advocates the design of detection algorithms to maximize the probability of detection for a fixed probability of false alarm) is not a measure of the reliability in the individual case, because consideration about samples which have *not* been observed, are simply not relevant to the problem of how we should reason from the one that *has* been observed.

Regarding the practical use of the proposed detection algorithm, 1200 frames of data currently requires about 55 s to process on a Pentium II class personal computer with a clock speed of 400 MHz. This is roughly a computational load of about 0.05 s per data frame, implying that the algorithm can be applied to particle count data in real time as it is acquired. This algorithm has be implemented in 4WARN, and uses a continuously updated background queue (containing usually 4 or 5 frames of background count data for a sampling time of 12 or 15 s) to store the current background sample to be used as the reference for the subsequent detection decision of the next frame of received data. Although the background rate b is assumed to be constant in the sample contained in the background queue, it is important to emphasize that this rate can change from one update of the queue to the next. Hence, non-stationary background rates can be accomodated by the proposed detection algorithm. All that is required by the detection algorithm is that the background rate in the given background sample remain constant, which does not disallow the case where the background rate is changing from sample to sample.

REFERENCES

1. Birenzvige, A., "Respirable Particles' Concentration as an Indicator of a Biological Attack (U)", ERDEC–TR–532, Edgewood Research, Development, and Engineering Center, Aberdeen Proving Ground, MD, 1998 (Secret).

2. Helström, C. W., "Statistical Theory of Signal Detection", Pergamon, Elmsford, New York, 1968.

3. Jeffreys, H., "Theory of Probability", Oxford University Press, Oxford, 1961.

4. Parzen, E., "Modern Probability Theory and its Application", John Wiley & Sons, Inc., New York, 1960.

5. Jaynes, E. T., "Prior Probabilities", IEEE Trans. Systems Sci. Cybern., SSC-4, pp. 227–241, 1968.

6. Laplace, P. S., "A Philosophical Essay on Probabilities", unabridged and unaltered reprint of Truscott and Emory translation, Dover Publications Inc., New York, 1951 (original publication date 1814).

7. Spanier, J. and Oldham, K. B., "An Atlas of Functions", Hemisphere Publishing Corporation, New York, 1987.

8. Erdélyi, A. (Editor), "Tables of Integral Transforms: Volume II", McGraw-Hill Book Company, Inc., New York, 1954.

9. Hairston, P. P., Ho, J., and Quant F. R., "Design of an Instrument for Real-Time Detection of Bioaerosols Using Simultaneous Measurement of Particle Aerodynamic Size and Intrinsic Fluorescence", J. Aerosol Sci., 28, pp. 471–482, 1997.

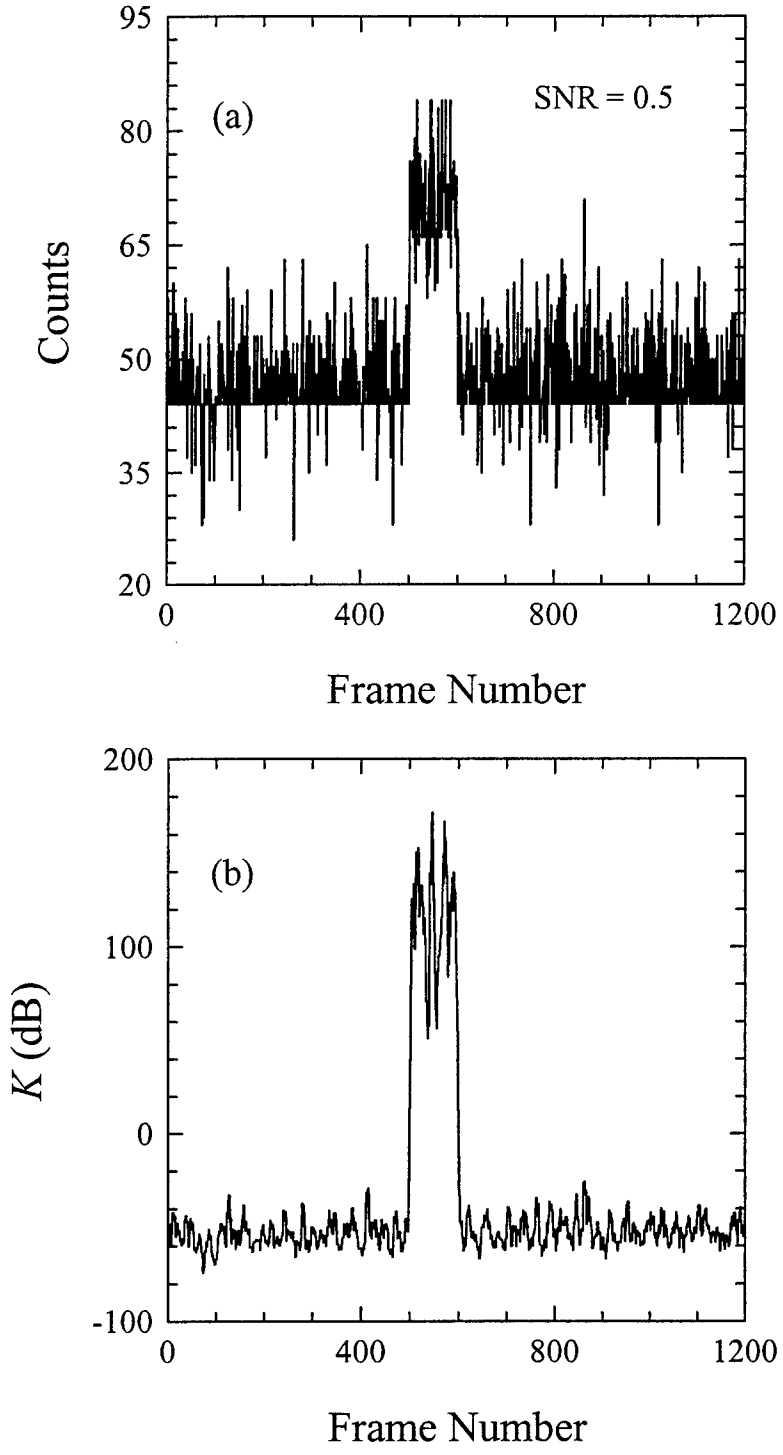

Figure 1. (a) Simulated data for signal particle counts at SNR = 0.5. (b) Evidence K for presence of a bio-target signal for data in (a).

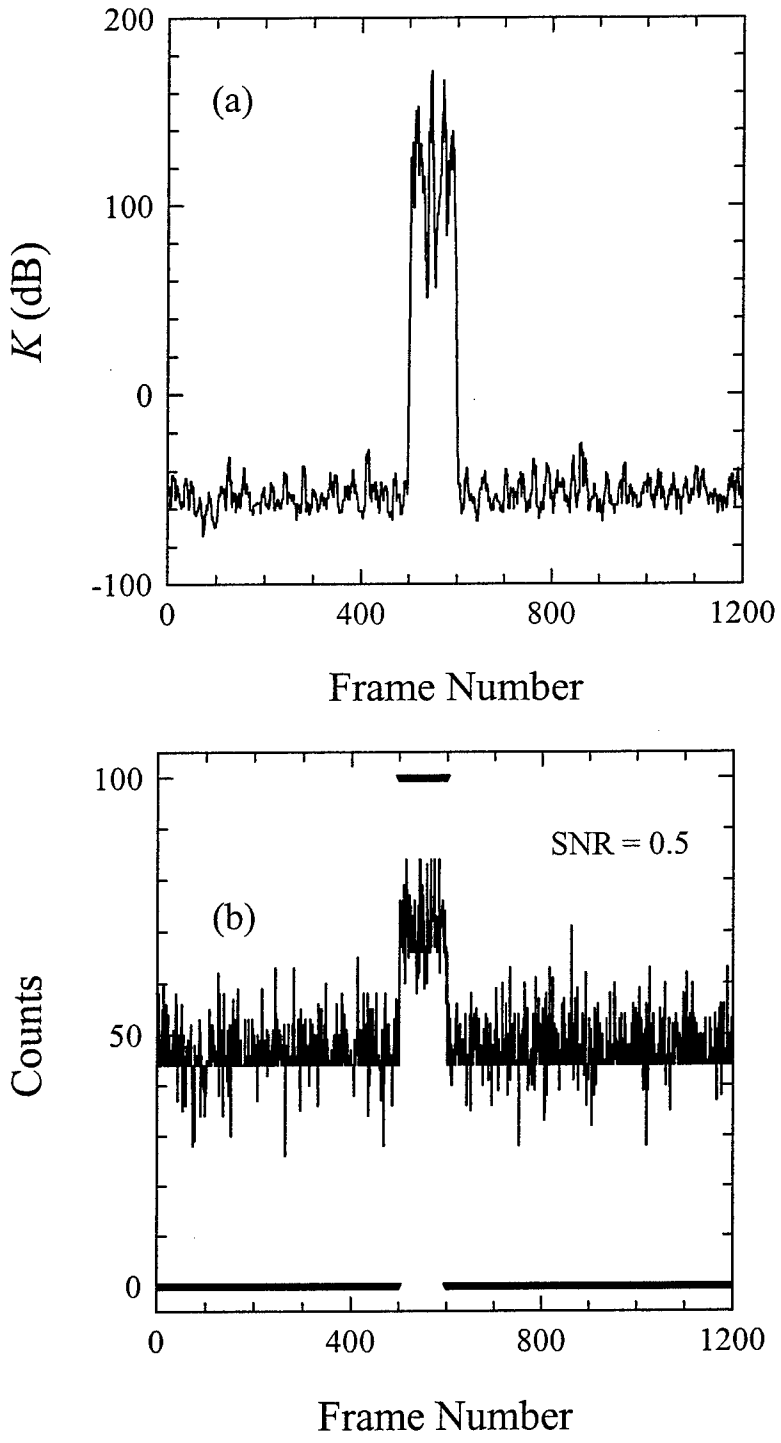

Figure 2. (a) Evidence K for detection of bio-target signal shown in Figure 1(a). (b) Indicator function (shown by inverted triangle) for detection of signal (threshold of 0 dB used on evidence).

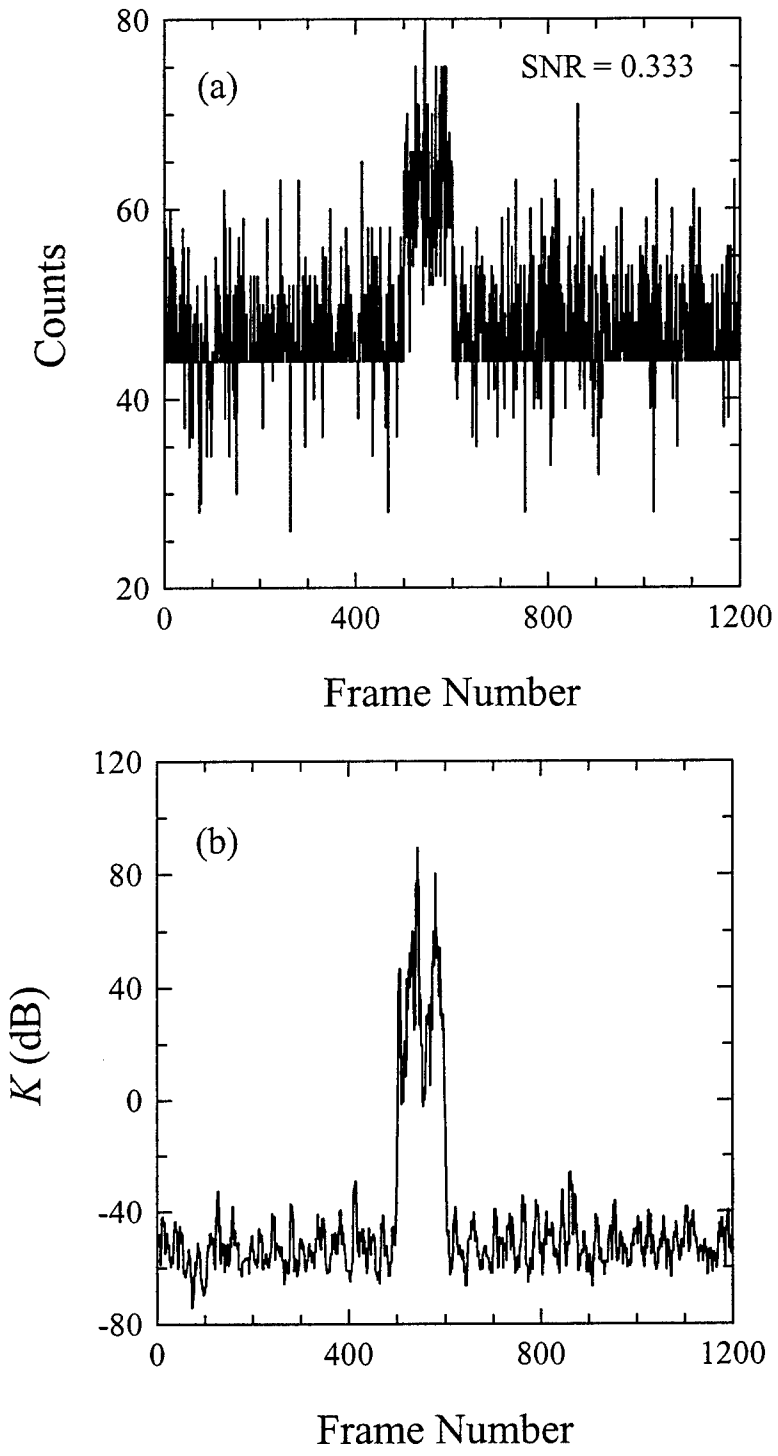

Figure 3. (a) Simulated data for signal particle counts at SNR = 0.333. (b) Evidence K for presence of a bio-target signal for data in (a).

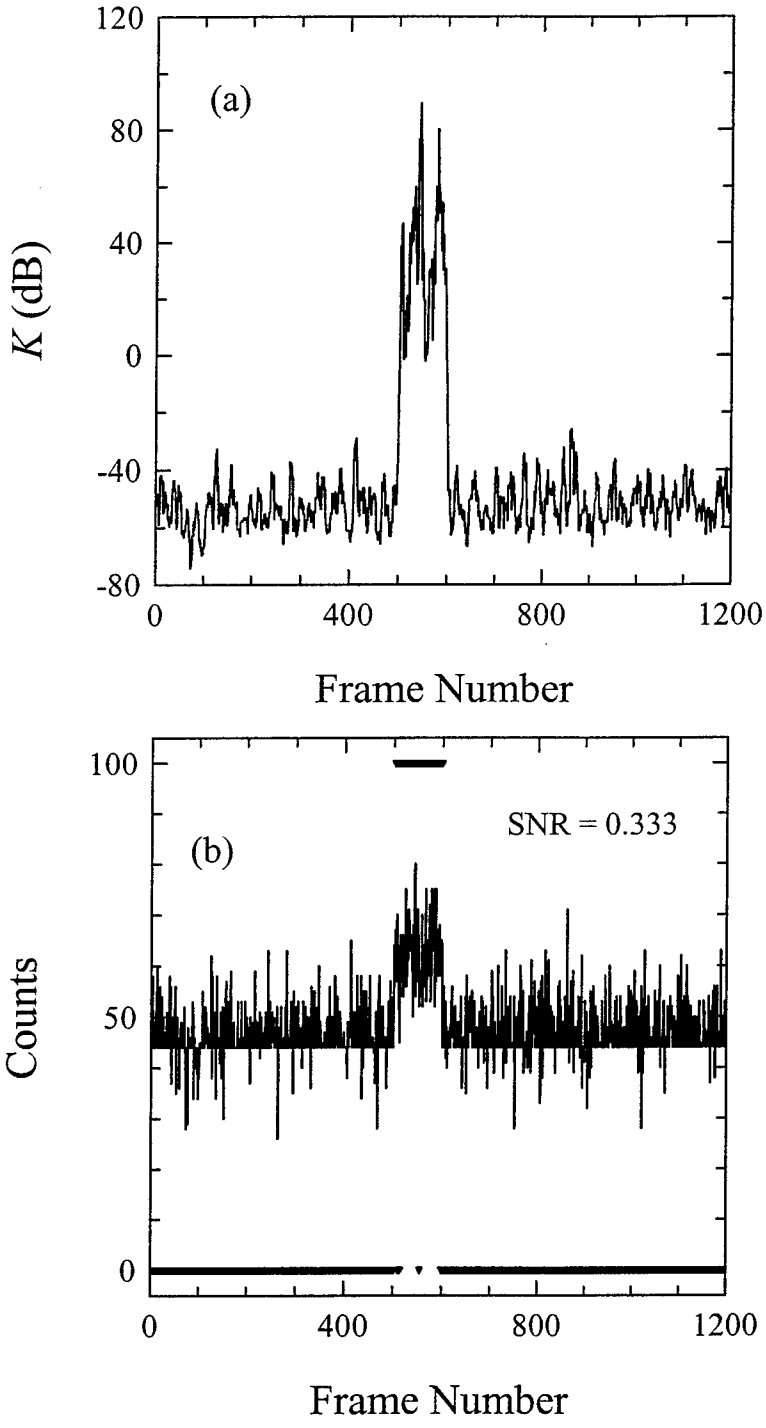

Figure 4. (a) Evidence K for detection of bio-target signal shown in Figure 3(a). (b) Indicator function (shown by inverted triangle) for detection of signal (threshold of 0 dB used on evidence).

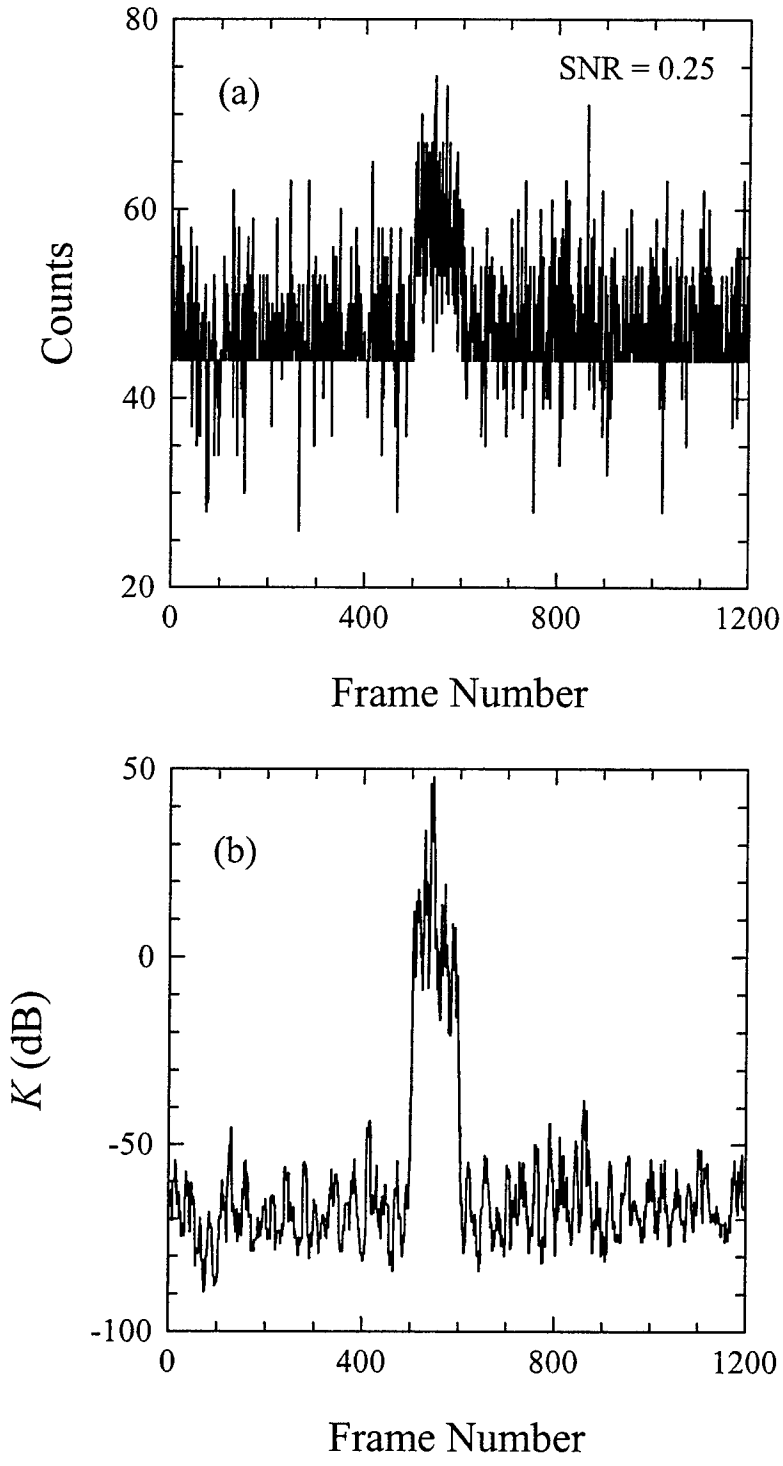

Figure 5. (a) Simulated data for signal particle counts at SNR = 0.25. (b) Evidence K for presence of a bio-target signal for data in (a).

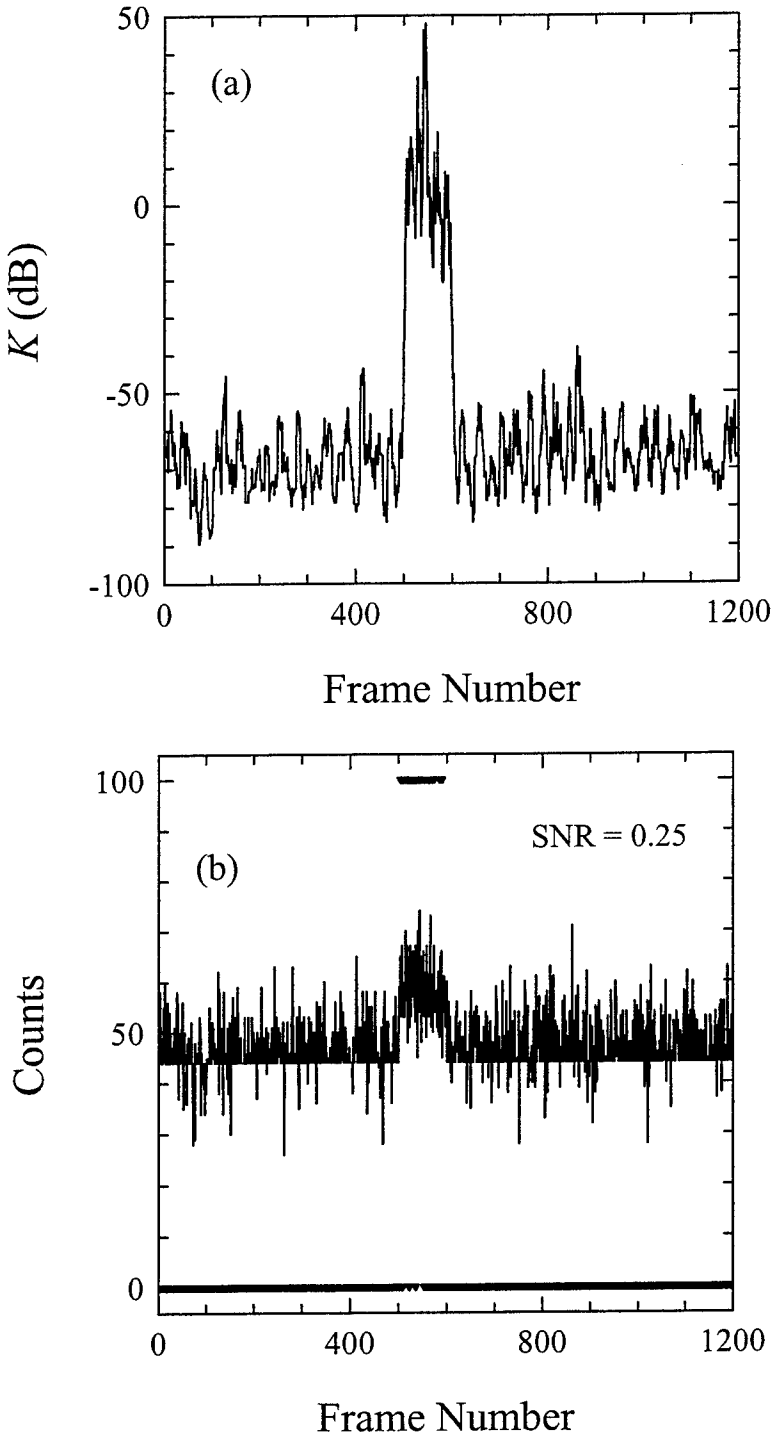

Figure 6. (a) Evidence K for detection of bio-target signal shown in Figure 5(a). (b) Indicator function (shown by inverted triangle) for detection of signal (threshold of 0 dB used on evidence).

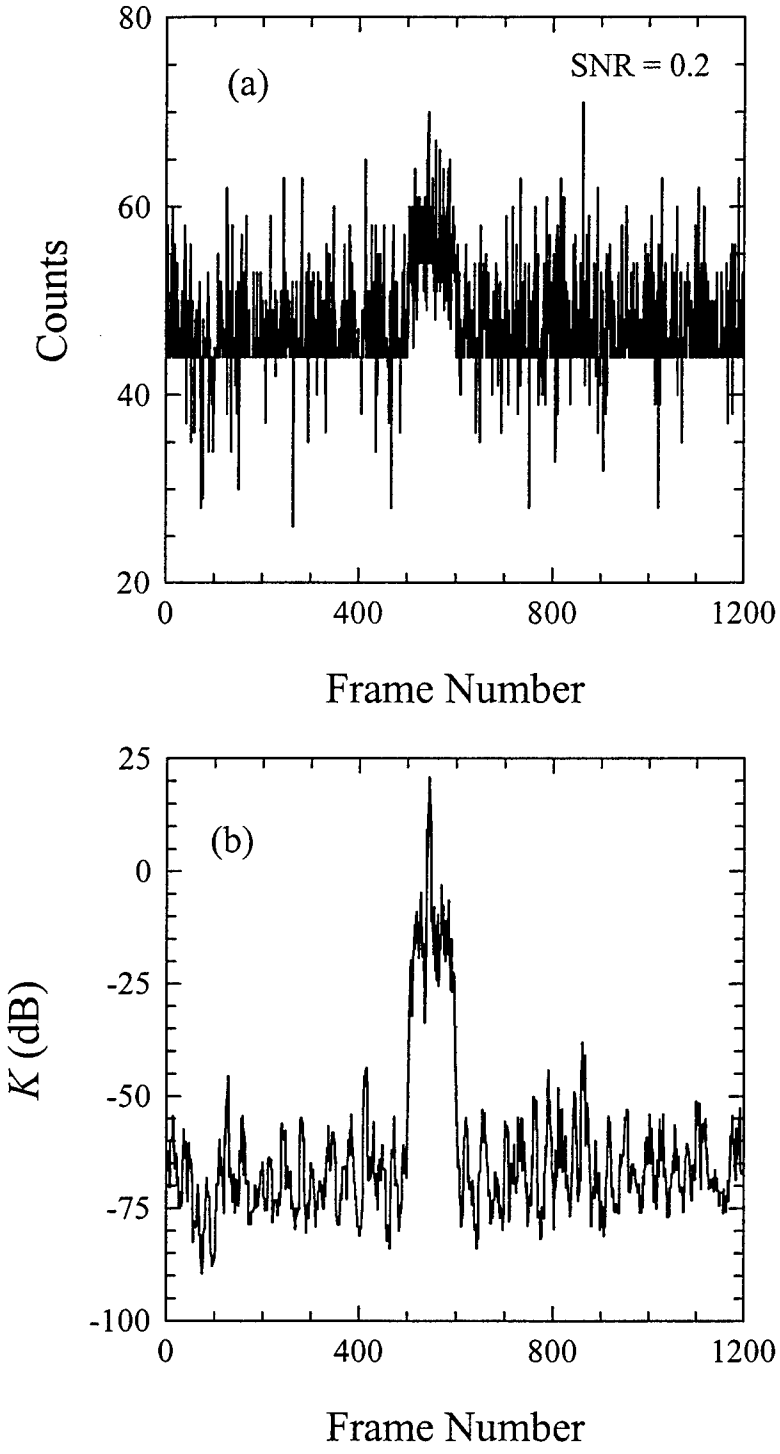

Figure 7. (a) Simulated data for signal particle counts at SNR = 0.2. (b) Evidence K for presence of a bio-target signal for data in (a).

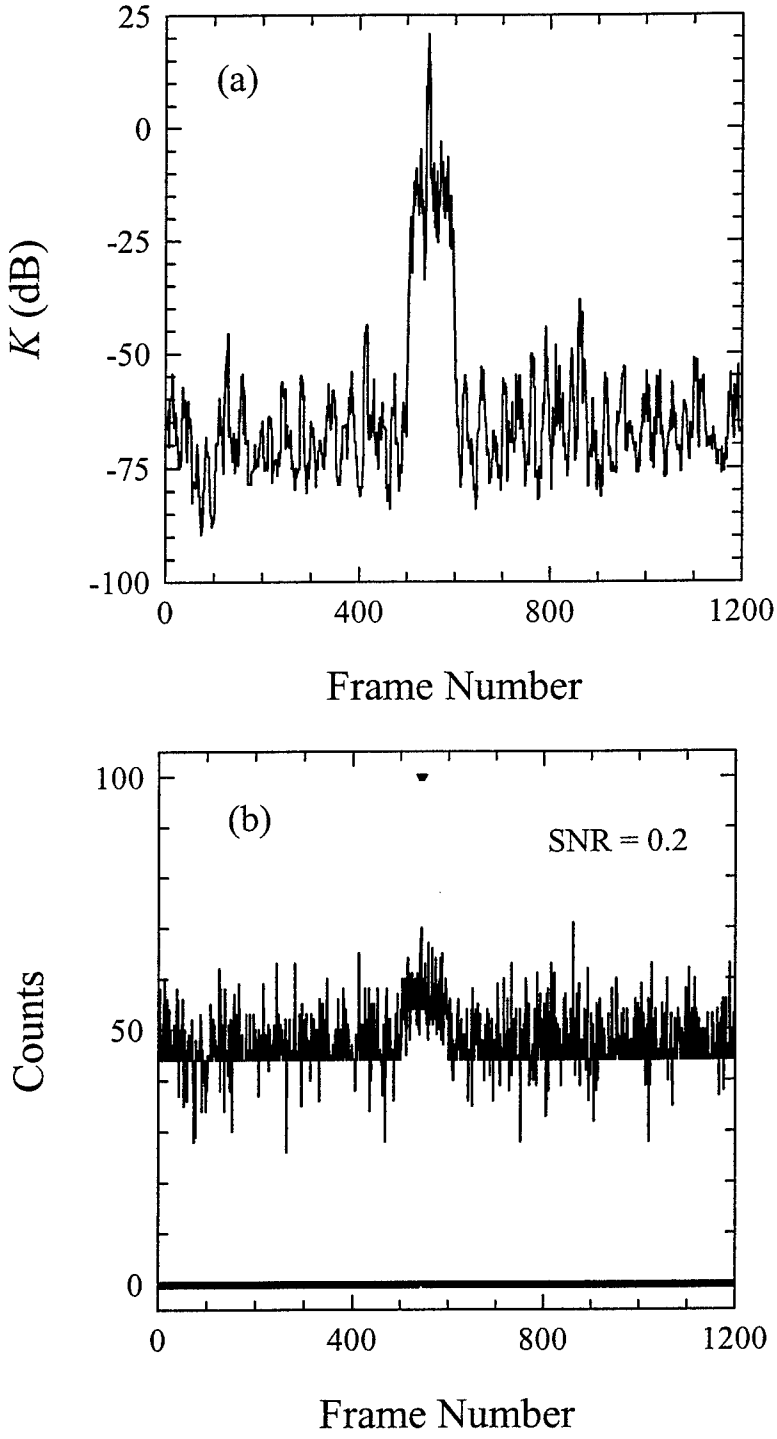

Figure 8. (a) Evidence K for detection of bio-target signal shown in Figure 7(a). (b) Indicator function (shown by inverted triangle) for detection of signal (threshold of 0 dB used on evidence).

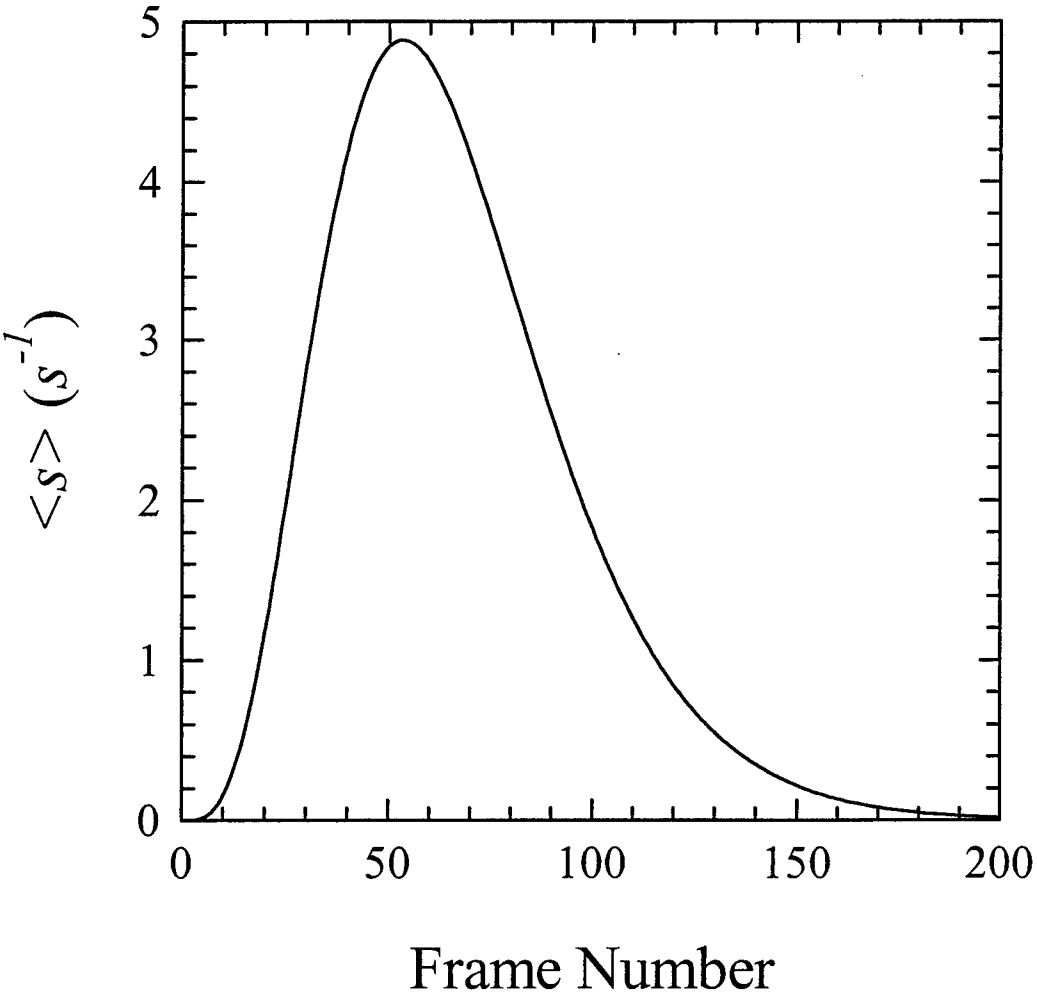

Figure 9. Expected (mean) bio-target signal rate for a transient, time-varying signal.

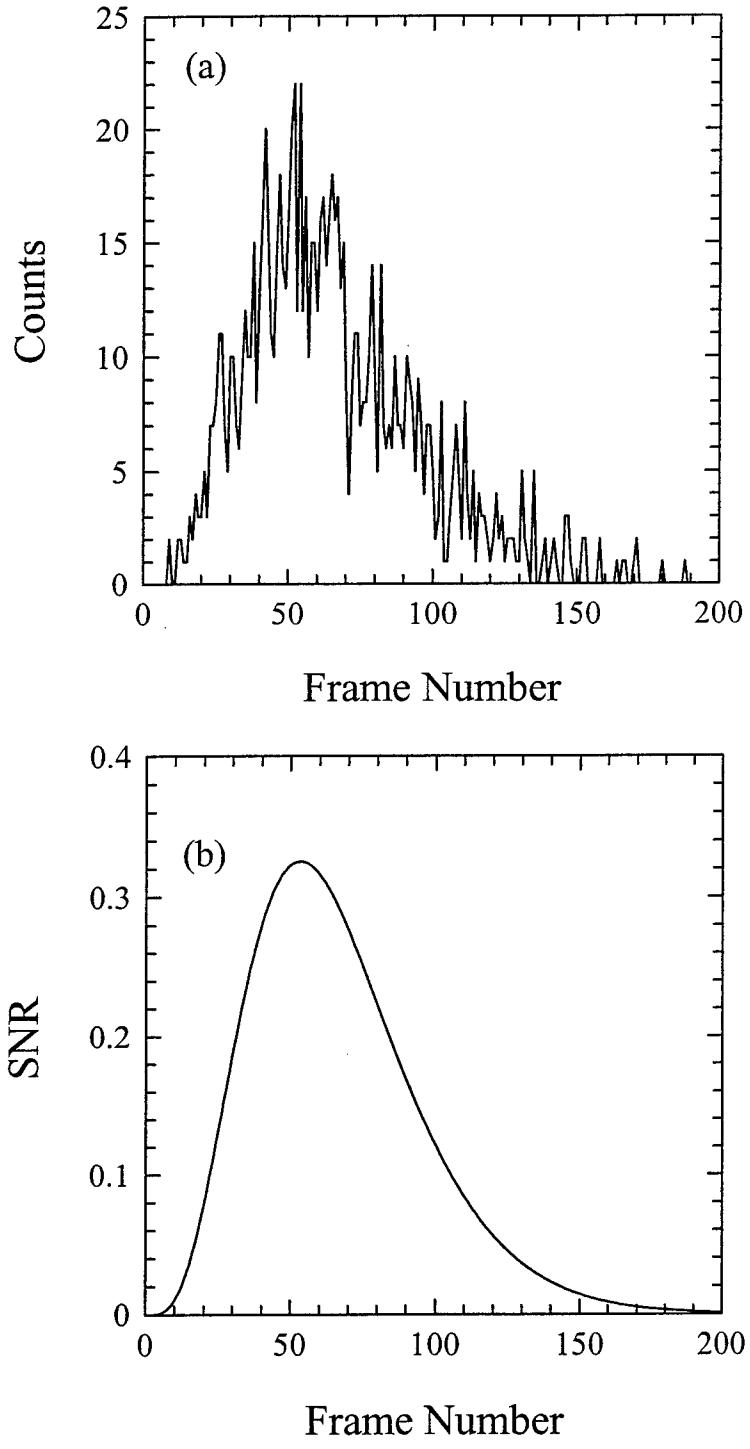

Figure 10. (a) A realization of bio-target signal counts for the expected signal rate shown in Figure 9. (b) Signal-to-noise ratio of the bio-target signal when added to a constant background interference with an expected background rate of 15.0 s^{-1}.

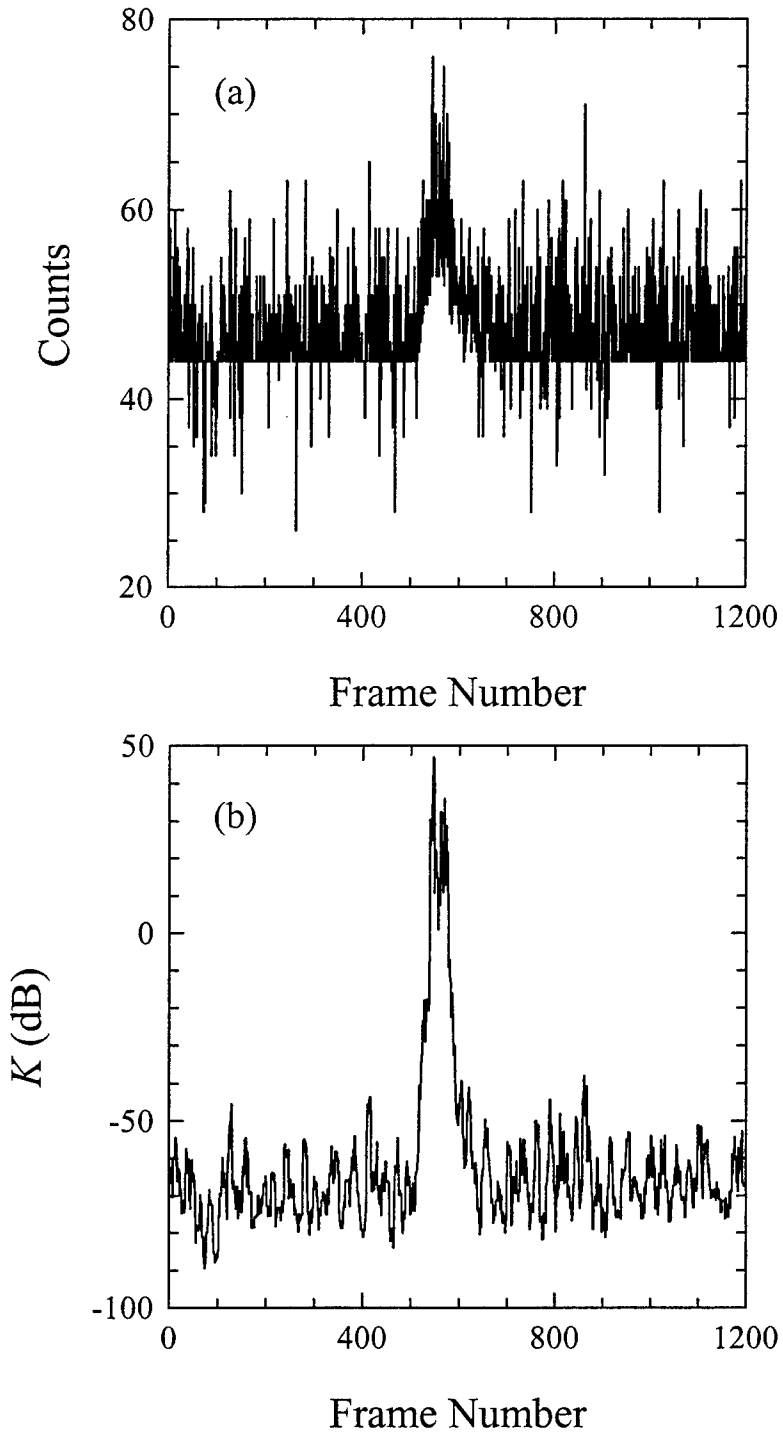

Figure 11. (a) Simulated data for transient signal particle counts with the time-varying expected rate shown in Figure 9. (b) Evidence K for presence of a bio-target signal for data in (a).

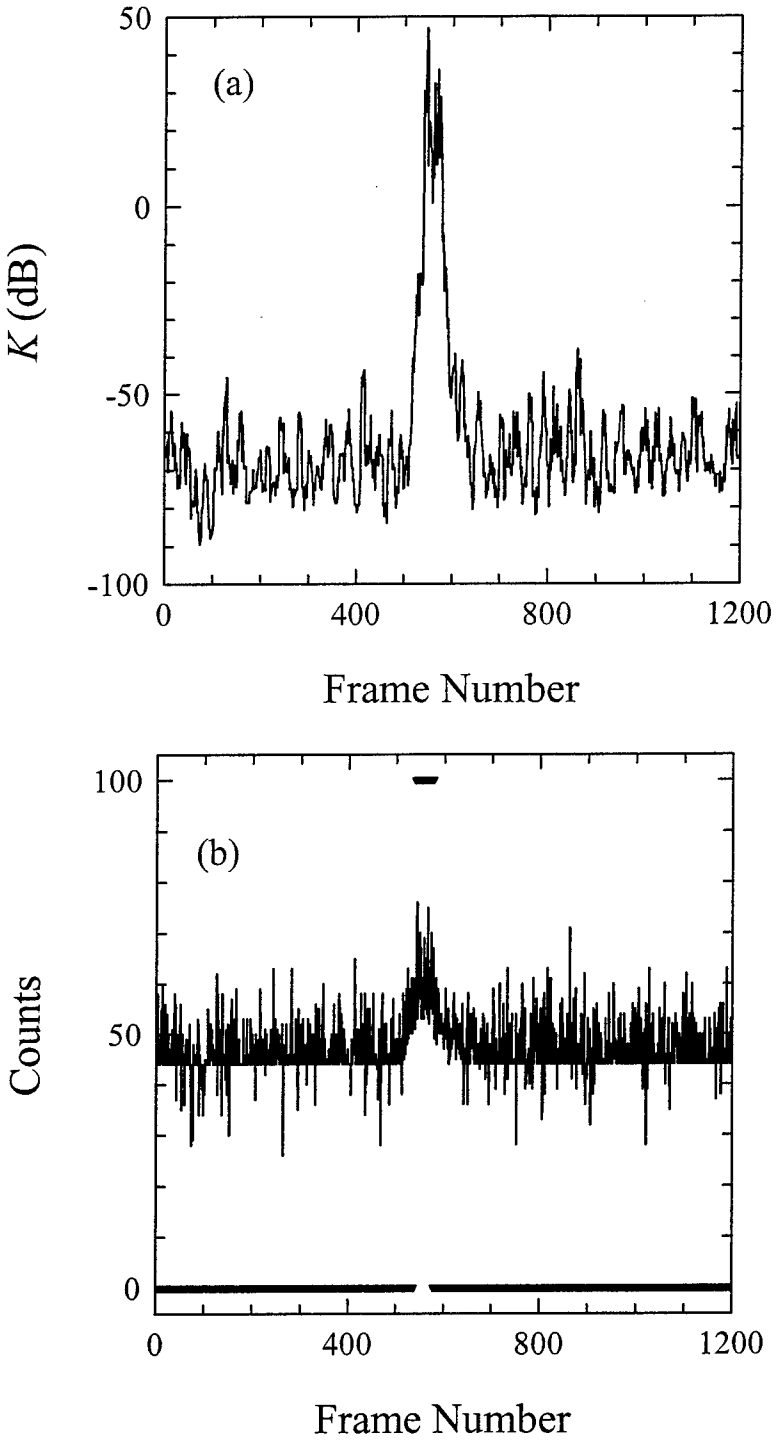

Figure 12. (a) Evidence *K* for detection of bio-target signal shown in Figure 11(a). (b) Indicator function (shown by inverted triangle) for detection of signal (threshold of 0 dB used on evidence).

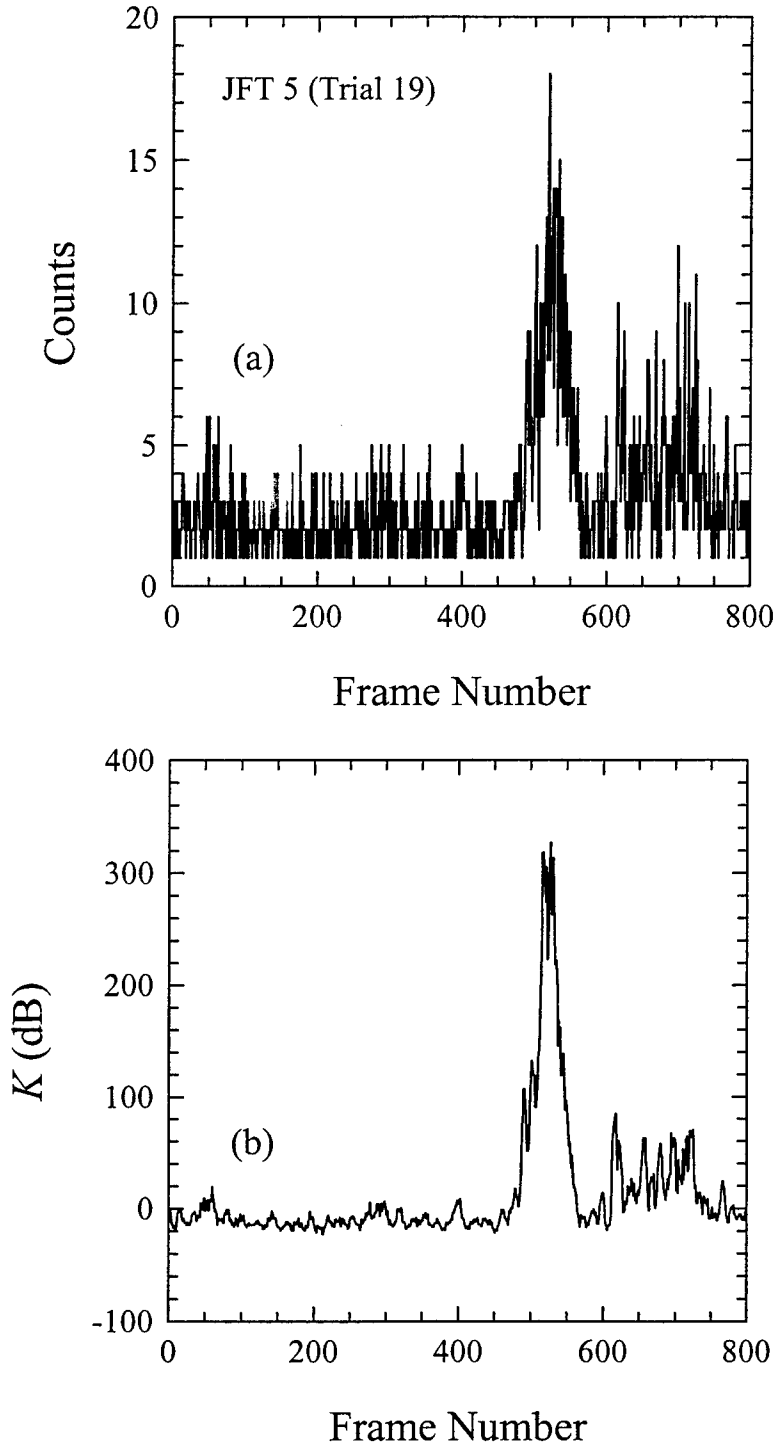

Figure 13. (a) Particle count data measured in a prescribed window of particle size and fluorescence intensity in Trial 19 of JFT 5. (b) Evidence *K* for presence of bio-target signal for data in (a).

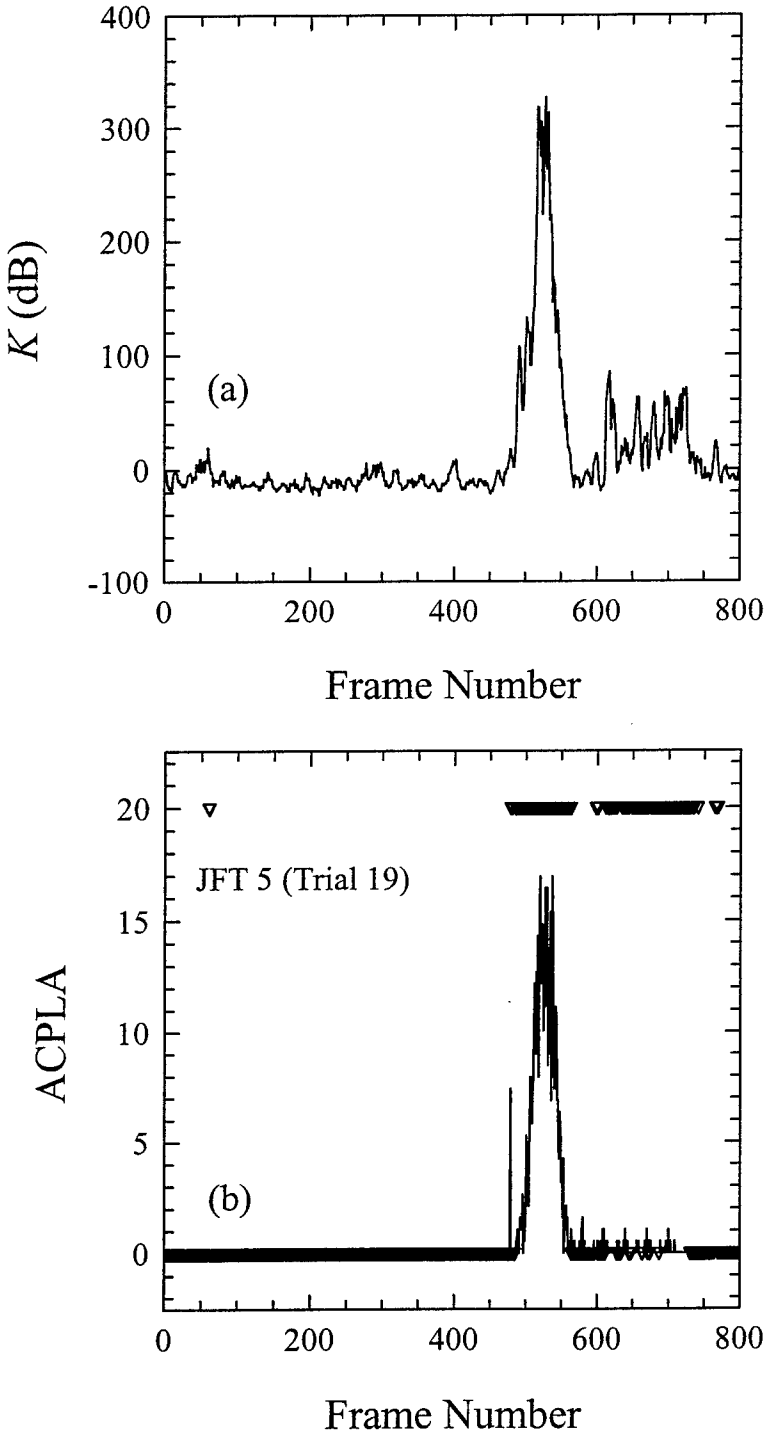

Figure 14. (a) Evidence K for detection of bio-target signal shown in Figure 13(a). (b) Indicator function (shown by inverted triangle) for detection of signal (threshold of 7 dB used on evidence).

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Establishment sponsoring a contractor's report, or tasking agency, are entered in section 8.)

Defence Research Establishment Suffield
P.O. Box 4000
Medicine Hat, Alberta T1A 8K6

2. SECURITY CLASSIFICATION (overall security classification of the document including special warning terms if applicable)

UNCLASSIFIED

3. TITLE (the complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S,C,R or U) in parentheses after the title.)

On the Theory of Particle Count Detection with an Application to the Triggering of Biological Warfare Detection Systems (U)

4. AUTHORS (Last name, first name, middle initial. If military, show rank, e.g. Doe, Maj. John E.)

Yee, Eugene C.

5. DATE OF PUBLICATION (month and year of publication of document)

March 2000

6a. NO. OF PAGES (total containing information. Include Annexes, Appendices, etc.)

44

6b. NO. OF REFS (total cited in document)

9

6. DESCRIPTIVE NOTES (the category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Technical Memorandum (Final) [1 January, 1999 to 31 March, 1999]

8. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include the address.)

Defence Research Establishment Suffield
P.O. Box 4000
Medicine Hat, Alberta T1A 8K6

9a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant)

PCN No. 6QD11

9b. CONTRACT NO. (if appropriate, the applicable number under which the document was written)

10a. ORIGINATOR'S DOCUMENT NUMBER (the official document number by which the document is identified by the originating activity. This number must be unique to this document.)

DRES TM 1999-062

10b. OTHER DOCUMENT NOS. (Any other numbers which may be assigned this document either by the originator or by the sponsor)

11. DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification)

(✓) Unlimited distribution
( ) Distribution limited to defence departments and defence contractors; further distribution only as approved
( ) Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved
( ) Distribution limited to government departments and agencies; further distribution only as approved
( ) Distribution limited to defence departments; further distribution only as approved
( ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availabilty (11). However, where further distribution (beyond the audience specified in 11) is possible, a wider announcement audience may be selected.)

Unlimited

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both offical languages unless the text is bilingual).

A new procedure is presented for the detection of a bio-target signal in aerosol particle number count data when no prior knowledge of the existence of such a signal or of its characteristics (e.g., amplitude and shape) is assumed. Unlike previous bio-target detection algorithms, the algorithm in this paper is derived rigorously by the direct application of probability theory. To address the detection problem, probability theory is used to compare two models (or hypotheses); namely, a model (M_1) that postulates the presence of a bio-target signal in the background interference, and an alternative model (M_2) that postulates the presence of a bio-target signal in the background interference. The posterior probability for each model is calculated based on all the available prior information, and used to determine the posterior odds ratio O_{12} in favor of model M_2 over model M_1. The ratio provides a quantitative measure of the evidence for the presence of a bio-target signal in the data. The new detection algorithm has been applied to both simulated and real particle count data and found to perform well.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

Biological Warfare Detection
Probability Theory
Particle Counting

The Defence Research
and Development Branch
provides Science and
Technology leadership
in the advancement and
maintenance of Canada's
defence capabilities.

Leader en sciences et
technologie de la défense,
la Direction de la recherche
et du développement pour
la défense contribue
à maintenir et à
accroître les compétences
du Canada dans
ce domaine.

DEFENCE R&D DÉFENSE

www.crad.dnd.ca